DOCUMENT RESUME

ED 124 869 CS 002 581

AUTHOR Wearing, Alexander
TITLE Memory for Sentences.
INSTITUTION Illinois Univ., Urbana. Computer-Based Education Lab.
SPONS AGENCY Advanced Research Projects Agency (DOD), Washington, D.C.; National Science Foundation, Washington, D.C.; Public Health Service (DHEW), Arlington, Va. Div. of Nursing.
REPORT NO CERL-X-6
PUB DATE Apr 69
GRANT NSF-GJ-81; ONR-Nonr-3985(08)
NOTE 181p.; Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Grammar; Language Research; *Memory; Psychological Studies; Recall (Psychological); *Retention Studies; Semantics; *Sentences; *Sentence Structure; *Syntax; Transformation Generative Grammar

ABSTRACT

Two experiments were carried out to see what parts of a sentence are best retained, whether those parts retained best immediately after acquisition are also remembered best after a 48 hour retention interval, and whether or not the parts of a sentence which are retained depend on the semantic and syntactic structure of that sentence. In the first experiment, retention was assessed by a recognition method, and in the second experiment, it was measured by cued or primed recall. In the first experiment, high predictability and simple sentences resulted in significantly better recognition of old sentences at both the zero and 48 hour retention intervals. Transformational complexity was not significant, active and passive sentences being recognized equally well. The detection of word changes did not change over the 48 hour retention interval. In the cued recall experiment, high predictability and simple sentences again resulted in significantly better recognition at both time intervals. Another result which paralleled one from the first experiment was the relatively high level of recall of simple unpredictable sentences compared with complex unpredictable ones. The relationship of the findings to various theories of memory for sentences is discussed. (MKM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

726

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED124869

CERL REPORT X-6 APRIL 1969

MEMORY FOR SENTENCES

ALEXANDER WEARING

Computer-based Education Research Laboratory

University of Illinois Urbana Illinois

2

CS002581

BEST AVAILABLE COPY

This work was supported in part by the National Science Foundation under grant NSF GJ 81; in part by the Advanced Research Projects Agency under grant ONR Nonr 3985(08); in part by Project Grant NPC-188 under the Nurse Training Act of 1964, Division of Nursing, Public Health Service, U.S. Dept. of Health, Education and Welfare; and in part by the State of Illinois.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Distribution of this report is unlimited

MEMORY FOR SENTENCES

BY

ALEXANDER JAMES WEARING

B.A., University of Adelaide, 1963

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois, 1969

Urbana, Illinois

ABSTRACT

Two experiments were carried out to see (a) what parts of a sentence are best retained, (b) whether those parts retained best immediately after acquisition are also remembered best after a 48 hour retention interval, and (c) whether or not the parts of a sentence which are retained depend on the semantic and syntactic structure of that sentence.

In the first experiment, 72 Ss learned 64 naturally occurring sentences that were equally distributed across 8 categories: Active or Passive, Simple or Complex (with regard to degree of left-branching), and Predictable or Unpredictable. At the retention test, either immediately or 48 hours later, the Ss were presented with a test list including the 64 old sentences and 64 new sentences. The new sentences were formed by altering the old sentences in one of eight ways: either by shifting the adverbial phrase to an adjacent, intermediate or extreme position in the sentence, or the adverb (following the main verb) to the front of the sentence, or replacing the subject, object, verb, or adverbial phrase noun by a synonym or a near synonym. Each sentence was presented by itself, and S had to judge whether or not it was "old" or "new".

In the second experiment, 48 Ss learned the same 64 sentences learned in experiment 1. For the retention test the Ss were presented with two cue words (either the subject + adjective, verb + adverb, object + adjective, or adverbial phrase noun + adjective) from each sentence and asked to generate the rest of the sentence.

ACKNOWLEDGEMENTS

A dissertation depends on many people for its success. I was most fortunate in my companions and advisors.

Fellow graduate students criticized the original proposal, which was improved by the insights of Thomas Duffy, John Limber, John McIntyre and Howard Thorsheim. I owe a particular debt to Clinton Walker, who gave generously of his time and special abilities to remove many infelicities of expression, errors of fact, and illogical arguments from various drafts.

The execution of the experiment was expedited by the staff at CERL, Carl Edward Webber, who helped with programming, and a peal of belles, Marcia Harms, Valerie Kussler, Barbara Larrick, Pam Pohlman and Jill Roberts. Linda Kalmanek made a particularly constructive contribution to the analysis of the data.

My dissertation committee, Jack Adams (chairman), Don Dulany, Harold Hake, William Montague, Ledyard Tucker and Klaus Witz were invariably helpful, and I gratefully acknowledge their assistance. I particularly wish to thank Dr. Adams; his careful, rigorous, and always helpful guidance greatly improved both the quality of this thesis and the research skills of its author.

Rosemary Wearing, while pursuing her own graduate studies, typed innumerable drafts, was primarily responsible for developing the experimental sentences, and (as a good companion and research associate) cheerfully endured many long hours hunched over a hot computer.

TABLE OF CONTENTS

Page

INTRODUCTION 1
    Measurement of sentences: Content and Meaning 7
    Measurement of sentences: Structure 12
        Approximation to English (AE) 12
        Sentence Predictability 15
    Grammatical models of sentence structure 27
        Finite - state grammar 27
        Phrase - structure grammar (PSG) 29
        Transformational grammar (TG) 36
    Theories of memory and some related evidence 43
    Summary 64

THE PRESENT STUDY 68
    Experiment I 70
        Acquisition sentences 70
        Distractor sentences 72
        Experimental design 74
        Apparatus 75
        Procedure 76
        Results 78
    Experimental II 103
        Acquisition sentences 105
        Priming cues 105
        Experimental design 105
        Apparatus 105
        Procedure 105
        Results 106

DISCUSSION 117
    Effect of sentence type on retention 117
        Recognition 117
        Cued recall 118
    Theories of memory of sentences 119
        Surrogate structure theory 119
        Mnemic abstraction theory 120
        Deep structure theory 121
        Surface structure theory 123
    Some problems and theoretical suggestions 124

REFERENCES 129

APPENDICES 144

VITA 174

INTRODUCTION

Word strings have played only a small part in retention studies. There are at least two reasons for this. First, it has proved difficult to find adequate descriptions and measures of word strings and sentences. Second, association theories have always dominated verbal learning and these have directed the attention of experimenters towards simple connections between elementary events, and away from patterned relationships.

The particular problem in defining sentences is that whereas a single item (as in a list of nonsense syllables) can usually be described without reference to any other item, in sentences items are not independent. The role and meaning of a word depends on the sequence of words in which it is embedded, and so any analysis must take account not only of the individual items, but also of the various phrases (or structures) in which it participates.

The influence of association theory on the choice of stimulus materials for experiments is easily demonstrated. Association theories have generally held that all mental functioning can be reduced to elementary (neural) events which are concatenated in a linear fashion. That is, the elements themselves remain unchanged regardless of any combination into which they enter, and changes in the associative relationship between one pair of elements do not result in changes for any other combination of elements.

The problem of studying these combinations of mental elements led to the introduction of nonsense syllables by Ebbinghaus

(1885) because he believed they were uniform in learning difficulty, and lacked any previously established connection with one another. He expected that the referents of these nonsense syllables in the mind would be treated as separate and independent elements and their retrieval during a retention test could be presumed to reflect the associative strength of the particular item recalled, which itself was assumed to be a direct function of the amount of practice during learning.

The model of the mind implied by Ebbinhaus' (1885) approach is one of a network of associations between elements; the association being contingent on the temporally contiguous occurrence in the external world of the events represented by these elements; their strength depending on the frequency of contiguous occurrences. Since no internal structure was imputed to the mind, there was no need for the stimulus materials to possess inter-item structural relationships neither did they.

The influence of association theory has been pervasive. Stimulus materials have usually been single elements, like words or CVCs (Adams, 1967). The great interest in associative networks (Deese, 1962, 1965; Marshall & Cofer, 1965; Russell & Jenkins, 1954) further attests to its importance. One of the reasons for the popularity of the paired-associates (PA) learning technique is its presumed virtue of allowing lists of items to be treated as independent S-R elements (Battig, 1966; Underwood, 1965a). At times the disregard of structure has been carried to extremes: studies which have extended interference theory (a member of the family of association theories) to sentences have usually treated the sentences as undifferentiated item strings, merely equivalent to a list of homogeneous words, horizontally arranged (Hall, 1955; Slamecka, 1959, 1961, 1962; Slamecka & Ceraso, 1960).

It should not be supposed that association theories have either completely dominated or restrictively blinkered verbal learning. Indeed, in recent years, verbal learning experimenters have been fiercely a-theoretical. For many of them theoretical bases are too often theoretical biases; servants of their data, they have put their trust only in experimental results. Avoiding larger theoretical issues they have concentrated instead on investigating the functional relationships between variables, remaining content with relatively low order theoretical constructs and empirical generalizations.

However, the data themselves have led to problems of structure and the outcome of this willingness to follow the data has been an awakening of interest in structural relationships. These relationships may involve sequential structures (as in linguistic habits) or simultaneous structures, as in the organization of elements (ideas, images, r_ms , or whatever) in memory. We turn now to some of the verbal learning findings which suggest that sentences and word strings may be valuable for studying psychological processes like learning and remembering.

Despite the inherent difficulties for association theory in providing an account for complex sequential processes, verbal learning investigators have been traditionally interested in serial learning from the time of Ebbinghaus forward (Osgood, 1953, Pp. 502 ff; Woodworth & Schlosberg, 1954, Pp. 708 ff; Young, 1968). Association theories of serial processes have been developed (Hull, 1930; Slamecka, 1964a: Staats & Staats, 1966; Young, 1968), but the theories have never been effectively extended to problems of sentences and connected discourse. Nevertheless, it has always been intended that these models will eventually encompass any kind of verbal material (Staats & Staats, 1966), be it nonsense syllables or the New York Times. Certainly, the study of sentences is a logical extension of a traditional verbal learning interest in serial processes (Jenkins, 1965).

A phenomenon which has been identified in the verbal learning laboratory, and which seems particularly suited for studying with word strings, is response integration and organization. Underwood & Schulz (1960) have found that in PA learning it is necessary to integrate the response term into a unitary entity before it can be hooked up with the stimulus. Mandler (1954, 1962) notes that previously discrete parts of a response sequence can become integrated so that they behave as a functional unit. He shows that these integrated response sequences which are now independent of particular associative connections (e.g. as in learning to learn) may actually have arisen from simple associative processes. Battig (1966, 1968), Mandler (1967), and Tulving (1962, 1964, 1968) have all pointed to the importance of response organization in almost any form of learning, even within apparently pure association paradigms. Sentences are ideal for studying reponse integration as they possess structure on several levels; letters must be integrated into words, words into phrases, and finally the sentence itself must be integrated from the phrase structures.

In any case, it is difficult to escape the effects of a person's language repertoire and Underwood (1965b) has stressed the importance of linguistic habits in verbal learning. These can be introduced by the S, who may form natural language mediators between items in a PA learning task (Montague & Wearing, 1967; Montague, Adams & Kiess, 1966). It is doubtful whether it is possible to stop Ss from using mediators of that kind. They can also be introduced by the E, and the importance of experimentally stimulated linguistic habits in PA learning has also been shown by Rohwer & his associates (Rohwer, 1966; Rohwer & Lynch, 1966; Rohwer, Shuell, & Levin, 1967), who found that learning was facilitated when the two items to be learned were embedded in a sentence. Any attempt to provide a satisfactory account of verbal learning must eventually deal with linguistic processes. It may even be argued that sentences are superior to single items in some respects as materials for verbal learning experiments in that they may reduce inter-subject variability in the kind of intellectual operations applied to the task, and allow the experimenter to exercise more control over the stimulus environment.

Since it seems that the study of sentences is a natural outgrowth of traditional interests in verbal learning, and that some problems are perhaps best studied by means of word strings, why is it that sentences have been largely ignored? Certainly the influence of association theory, and the lack of structural measures provide two reasons. But maybe there is a third. Although few would deny that sentences are a good deal more important in the work-a-day world than nonsense syllables, many feel that before tackling such complex applied problems it is first better to gain a full understanding of the simpler world of single items. It is argued that only when CVCs, nonsense syllables and words are safely encaged in our theoretical nets should we venture out after the big game of sentences and connected discourse.

However, there is an important qualitative difference between sentences and single items in that the former possess inter-item structure. Because of this difference, there is a point at which the further study of single items can yield no more information about sentences. If we are to increase our understanding, there is then no choice but to grasp the complex nettle of connected discourse with whatever tools are available.

The strategy should be to use what descriptions we can find for sentences, investigate the learning and memory of these sentences when well understood variables (in traditional verbal learning situations) like degree of practice and length of the retention interval are manipulated, compare the results with those from other verbal learning studies, and use these comparisons as a basis for further investigations.

The task of this investigation is three-fold; first, to review the sentence descriptions in the literature, and select those which seem to deserve further attention: second, to see what kind of theoretical formulations might accommodate the data on the retention of sentences: third, to generate some data on sentence retention, being particularly concerned to note the effects on remembering of traditional variables, like length of retention interval and level of learning.

Measurement of sentences: Content and Meaning.

The major difficulties of measurement encountered in studying sentences and prose, (as opposed to single items), were identified by Henderson (1903), one of the first investigators in the area. He noted that the overall meaning of a sentence is something more than a straightforward combination of meanings of its constituent words. A simple word count will not do as an index of a variable such as learning difficulty, since some words are crucial to the meaning of a sentence, while others are redundant.

Henderson introduced the notion of the meaning unit or idea group (IG) as an entity which might serve as the constituent element of sentences and prose passages. The IG may be defined as a linked set of words which can be seen as an ideational unit. The excerpt below from Henderson shows how a sentence may be partitioned into IGs, the diagonal slashes marking the boundary of each IG.

> "/A bear, / climbing over the fence / into a yard/ where
> bees were kept. / began at once / to smash the hives, /
> and to rob them / of their honey./"

Henderson admitted that his divisions were subjective and somewhat arbitrary, but he argued that anyone else's partitioning would not be sufficiently different to alter materially his results. Despite these doubts, the IG has been widely used (Clark, 1940; Cofer, 1941; Levine & Murphy, 1941; Lyon, 1916, 1917). Welborn & English (1937), in a review of the studies on logical and rote memory, found that Henderson's measuring methods were still widely employed in essentially unchanged form.

But the arbitrary way in which IG boundaries have been defined has continued to worry students of word strings. Cofer (1941) was able to generate three sets of different IGs from the same passage. Kay (1955) concluded that the whole business of IG partitioning smacks a good deal more of art than science. However, the only systematic methodological study on IGs that has been reported is due to Levitt (1956). He carried out a series of experiments and found that (a) different judges are likely to divide the same material in different ways, (b) recall scores are sensitive to the way in which material is divided, and (c) recall is affected by variations in IG size.

Nevertheless, despite its problems, the IG has proved a useful measure of learning and retention. Henderson's (1903) claim that small changes in IG groupings would have little effect on the results has been generally borne out. Cofer (1941) found that several scores, including retention measured by three different IG classifications, total number of words remembered and the number of E-defined "significant" words recalled were all quite highly related. The question then arises as to what measures are best, and how many to use. On the one hand it seems as though IGs are not really satisfactory as a measure of learning and retention, and on the other hand there is a little evidence that almost any measure will do, even just a gross word count (Adams, 1967, p. 169).

One way of meeting the problem of which measure to use is to factor analyse a number of indices, and so find out how many independent scores are needed to define learning and retention. King (1960) required his S̲s̲ to learn either one of two prose passages (one being Bartlett's (1932) War of the Ghosts) and he scored their recall protocols in seven different ways. These were:

(1) number of IGs recalled, (2) number of sentences present in recall, (3)number of content words recalled, (4) total number of words recalled, (5) number of words recalled that were also present in the original passage, (6) overall goodness of recall (as rated by a set of judges), and (7) a modification of the Cloze procedure (Taylor, 1953, 1956), which consisted of deleting every tenth word from the original passage and then determing whether or not it could be filled in from the recall protocol.

King first inter-correlated and then factor analysed these variables and found two factors. One factor was defined by the number of content words and IGs recalled, as well as by the modified Cloze score, and the other factor was defined by the total number of words recalled, and the number of words in common between the original and the recalled passage. He interpreted the first as a content or organization factor and the second as a length or quantity factor. Subsequent work by King and his associates (King, 1961; King & Lau, 1963; King & Russell, 1966; King & Schultz, 1960; King & Yu, 1962) has substantially replicated these results.

Most of King's measures loaded on both of the factors, which implies that his scores probably consist of two components, a general learning component and either an organization or a quantity component. Any two measures will be correlated (in general) because they share a learning component in common. In some of his other studies there was a fair amount of variation in the loadings of the variables, but in almost every case a two factor solution has emerged, with one factor primarily related to content and the other to length.

These results suggest that two correlated scores are enough to describe performance on a retention test for sentences (although a single measure could serve). The first score is the total number of words remembered, and the second is the amount of substantive meaning remembered, as indexed by the number of content words or IGs recalled.

In the division of sentences into IGs experimenters recognize that not all words make the same contribution to the meaning of a passage. The partitioning is usually done by isolating phrases containing nouns, the boundaries between IGs being indicated by function words and punctuation marks. The assumption underlying this procedure is that meaning (or content or semantic) density is related to learning, and that meaning is carried by content words. There is no attempt to further define or measure meaning density; whatever it is, the number of IGs in a passage are presumed to reflect its value for that passage.

While the distinction between content and function words (although not explicitly developed by these investigators) is useful and important (Glanzer, 1962) there are two serious objections to IGs as a measure. The first is that adjectives, although content words, have a different status from nouns, which they qualify. An IG grouping places nouns and adjectives together and treat a noun phrase with an adjective the same as a noun phrase without one. Using only content words will not help, for with that measure an adjective is not distinguished from a noun.

The second problem is that IGs (and content word measures) cannot take account of sequential constraints, be they structured or semantic. Sentences may be more or less redundant, and surely a highly redundant sentence has less meaning, in some sense, than one which is not so. It certainly should be easier to learn. The problem of structural constraint is treated in the next section. The question of semantic constraint has been tackled by Becker & Carroll (1963), who have developed a coefficient of sentence contingency, I_p. In order to obtain I_p they first define two scores, $O_p\%$ (Overlap percentage), which is the number of repeated nouns and pronouns (appropriately weighted) of the total number of nouns, (expressed as a percentage) and $C_p\%$ (concept percentage), which is the total number of nouns less the numerator of $O_p\%$ (the number of repeated nouns and pronouns) divided by the total number of words (expressed as a percentage). I_p is obtained by simply dividing $O_p\%$ by $C_p\%$. Becker & Carroll claim that I_p is roughly equivalent to redundancy and they have found that the learnability of prose passages is a function of I_p.

While the studies discussed in this section have not provided any wholly satisfactory empirical measures of sentences, they have indicated some important variables. The distinction between content and function words, and the importance of meaning density, organizational factors and sequential constraints, have all been demonstrated, if no explicitly. The tasks remaining are to specify structure rather more precisely, and to discover how the semantic characteristics of a sentence interact with its structural components to affect learning and retention.

Measurement of sentences: Structure

Approximation to English (AE). The studies discussed so far have focussed mainly on meaning and content. A pioneering study by Miller & Selfridge (1950) took another tack. They began with the premise that one of the fundamental characteristics of language is its structural organization or patterning. Their assumption was not particularly controversial; Henderson (1903) said much the same thing. But instead of defining the pattern in terms of meaning, Miller & Selfridge defined the influence of a verbal pattern or context as the extent to which the choice of a particular word depends on the words that precede it.

In order to examine this proposition, Miller & Selfridge compared word strings varying in their degree of patterning. They generated passages of different orders of approximations to English (AE), where the order of AE was defined in terms of the length of the preceding contextual string of words that determined a particular word.

Thus zero order AE sequences were obtained by drawing random samples of words from the Thorndike-Lorge (1944) lists. First order AE required that the relative frequency of words in the natural language be represented; therefore words were drawn at random from conventional printed texts.

Higher order AE passages were generated by human Ss. For instance, second order AE was produced by presenting a common word such as he, she, or it to a person who was instructed to use the word in a sentence, for example, "It goes on faith and four wheels". The word directly following the first word that was given, goes, was then presented to another S; who had to use it in a sentence, for example, "No man goes easily to his death". The word easily would then be presented to a third S, who would have to put it in a sentence. This procedure was followed until a sequence of words of the desired length had been constructed. Each successive pair of words went together in a sentence. Each word was determined in the context of only one preceding word.

For AE orders of three through to seven (sixth order was omitted), Ss saw a sequence of two or more words (depending on the order) each time and used the sentences in a sequence. Hence n^{th} order AE was a word string in which any set of n successive words went together in a sentence, and in which any one word was determined in the context of the previous (n-1) words.

Miller & Selfridge gave word strings varying in both length and order of AE to several groups of Ss, used immediate recall, and scored the number of words correctly recalled. They found that recall increased with the order of AE and decreased with increasing list length. In general, the higher orders were recalled as well as organized text. They concluded that high AE material is easy to learn because it preserves the short range associations or sequential dependencies that are familiar to Ss, rather than because it is meaningful in any general sense. A number of studies have confirmed these findings (Postman & Adams, 1960; Richardson & Voss, 1960; Sharp, 1958).

Miller & Selfridge used the total number of words correctly remembered as their recall score. Marks & Jack (1952) and Coleman (1963) criticized this method as it fails to distinguish between words which are merely correct, and words which are correct in sequence. Now the short-range associations which are presumably preserved in high AE material consist of at least two factors. The first is inter-item associations of a semantic kind, and the second is syntax. Counting the number of words correct taps the effect of the first component, while the influence of changes in grammaticality is reflected in the sequences of correct words. Marks & Jack (1952) and Coleman (1963) found that when recall was scored in terms of correct sequences or words, recall improved across all orders of AE, instead of becoming asymptotic at about the 4th order, as Miller & Selfridge reported.

Tulving & Patkau (1962) interpreted sequences of correct words as chunks, following Miller (1956) and defined the chunk as:

> ".....any group of one or more items which occur in a subject's recall in the same sequence as the input list. For example, if the stimulus list consists of the sequence A,B,C,D,....V,W,X, and the subject recalls T,U,V,A,B,C,E,F, R,K, in this order, we assume that the ten items the subject recalled are organized into five adopted chunks: TUV, AB, DEF, R and K" (p.90).

Tulving & Patkau found that the number of words correctly recalled increased with the order of AE, but that the number of chunks remembered was invariant. Regardless of the order of AE, the mean number of chunks recalled was about 5 or 6.

These findings, which were confirmed by McNuhlty (1966) show that higher AE material is easier to learn because it can easily recoded, not because it preserves short-range associations per se. However, recoding utilizes both inter-item associations and grammatical constraint. Although these AE studies point to certain variables (short-range associations and sequential dependencies) and processes (recoding) which are important in learning word strings, AE itself is not a measure that can be used in the study of sentence retention for the simple reason that it is not a measure of sentences. In the next section we will look at ways in which the notion of sequential constraint can be applied to naturally occurring sentences.

Sentence Predictability. Approximation to English scores may also be interpreted as measures of sequential constraint since by definition the higher the order of AE, the longer the sequence which determines or constrains any particular word. Similarly, the higher the order of AE, the more predictable any particular word will be from the preceding context. Miller & Selfridge's technique allows the amount of sequential dependency to be specified precisely, but it carries the implication that naturally occurring sentences have a very high order of AE, and indeed, that the AE orders of all sentences are similar, since in every case the approximation to English is perfect. Yet intuitively, it is clear that prose passages, for example, vary a great deal in predictability. There is a world of difference between good journalism and James Joyce.

These differences in difficulty between example of English prose have been examined by investigators interested in predicting readability (Klare, 1952). The pioneer in the field is Flesch (1948) who combined average sentence length in words, average sentence length in syllables, and number of personal words into a formula of reading ease. Dale & Chall (1948) developed an index of reading difficulty that emphasized the importance of unfamiliar words. However, both these measures are based on the kinds of variables used by the writers who were discussed in the first section in that indices from individual elements, or simple word counts were used, and the structure between the words was ignored.

Two techniques have been developed which do provide some measure of sentence predictability. One, also intended as an index of readability, is the Cloze procedure (Taylor, 1953, 1956), in which every n^{th} word is struck from a passage, then the S has to guess what the deleted words are. The other method is to give the S the first word of a sentence, let him guess the second word, after which he is told the correct answer, then let him guess the third word, following which he is told what the word is and so on through to the end of the sentence. This method which we shall call the successive guessing technique, is rather akin to the anticipation method of paired associates (PA) learning.

Rubenstein & Aborn (1958) required Ss to learn prose passages and found that the Flesch, Dale-Chall, and successive guessing scores correlated .61, .75, and .73 respectively with the amount of learning. Since the correlation between Cloze and Dale-Chall scores is high, (.94 according to Taylor (1953)), it seems as though both learnability and predictability are strongly affected by word frequency.

It is obvious that predictability must depend on word frequency, since, in a statistical sense at least, a word is predictable simply because it is frequent. But predictability also depends on sequential constraints, and the importance of these have been documented by Miller & Selfridge (1950) and others. However, it may be the case that in natural language, sequential constraints are, for all practical purposes, the same, and the word frequency is the only wide ranging variable.

Aborn, Rubenstein & Sterling (1959) examined the predictability of words as a function of their form class (part of speech) and position in a sentence. Their sentences were of three lengths, and words were omitted from one of four possible positions, as shown in Table 1. The S's task was to fill in the blank space in each sentence. All words were classified as either nouns, verbs, adverbs, pronouns or function words.

They found that:

(1) form class, position of omission, and sentence length are all independently effective as sources of constraints on words in sentences, (2) the predictability of words in a class is in general related to the size of the class, (3) increasing the context beyond 11 words does not increase predictability, and (4) words in medial positions are more predictable than words in initial or final positions.

There are three studies which elaborate somewhat on these findings from Aborn et al (1959). Salzinger, Portnoy & Feldman (1962) administered Miller & Selfridge's (1950) passages with every 5th word deleted to Ss. Their task was to replace the missing words. They found that the number of words replaced which were in the correct grammatical category (even if themselves not actually correct) rose rapidly from zero order to third order AE, then flattened out, whereas the number of words correctly replaced rose more slowly, but continued to rise across the higher orders of AE. The authors concluded that Ss could use the syntactical information that was present in strings of three or four words

TABLE 1

Aborn *et al*'s experimental design.

	POSITION OF OMISSION			
SENTENCE LENGTH	First word	Early medial	Late medial	Last word
6 words	1	3	4	6
11 words	1	3	8	11
25 words	1	4	17	25

even though the context was still insufficient to determine the precise meaning of the missing word. This conclusion implies that *Ss* can use syntactic and semantic information separately in predicting what a deleted word might be.

Fillenbaum, Jones & Rapoport (1963) deleted either every second, third, fourth, fifth or sixth word from transcripts of running speech. They found that the predictability of the form class of words depended on the immediate context, but that prediction of specific words depended on the overall semantic context and the size of the form class of words from which the particular word was drawn. Again, these results suggest that grammatical and semantic information are subjectively separable.

Treisman (1965) attempted to measure the separate effects of semantic and grammatical constraints. She used 100 word passages: 1st, 2nd, 4th, 6th, 8th, and 9th orders of AE; two prose passages, one a highly predictable children's story about camping, the other an extract from the novel *Lord Jim* by Joseph Conrad; and a passage of "Syntactical English" in which the words were selected randomly from *Lord Jim*, except for the constraint that each word had to be the same part of speech as the word in the same position in a sample 100-word extract chosen to provide a grammatical skeleton. An example of the outcome of this procedure ~~was~~

> "Up that scene the way had forgotten, maddening lumpily down a beard. He is perfunctorily soft with them to scatter you if he was called and held to process....."

These kinds of word strings are also called anomalous sentences (Marks & Miller, 1964; Miller & Isard, 1963).

The experimental task was to guess every tenth word (previously deleted) so each *S* had to guess 10 words. There were two groups of 100 *Ss* each; each group had to guess a different lot of 10 words, so that *in toto* there were 100 guesses for each of the 20 words for all the 9 passages.

Treisman defined the information content of each word as $-\log_2 p_c$ where p_c is the probability that a word will be guessed correctly. An estimate of that probability was obtained from the frequency of correct guesses (each of the 20 deleted words had a total of 100 guesses). The correct word for any deletion was assigned a minimum probability of being guessed correctly of 0.10, hence the maximum value of the information content for any word was 6.64 bits. (The more often a word was guessed correctly the greater its redundancy or predictability, and so the smaller its information content). The average word information content (WI) for a passage was defined as the mean amount of information contained by the 20 deleted words.

Since Treisman felt that the competing responses in the total ensemble of *S*'s guesses might also affect his performance, the average amount of information contained by *all* the guesses made was calculated. If there are *n* different words guessed for a particular deletion, and if the frequency of guesses for the i^{th} word is p_i, then the entropy of the distribution of responses for each blank, or its distributional entropy (DE) may be defined as follows:

$$DE = - \qquad (p_i \log_2 p_i)$$

The DE of a particular passage was obtained by finding the mean DE for all its 20 deletions. An analysis of the conceptual constraints (or semantic context effects) was made by taking guessed words of similar meaning, grouping them in "synonym clusters", and calculating the DE of the clusters in the same way as for words.

To compute the grammatical constraints, the DE of parts of speech was calculated. For this calculation an S's guesses were classified as either nouns, pronouns, adjectives, adverbs, prepositions, conjunctions or articles. It should be noted that for both the DE of synonym clusters and of parts of speech, the maximum information content is less than for words, simply because the number of response alternatives has been reduced. For words, the possible number of different guesses was 100, but for parts of speech it was only 8.

Treisman's results are summarized in Figure 1. She found that increasing the order of AE increased the redundancy of all her measures. Although the easy prose was more redundant than 6th order AE, the difficult prose (the extract from Lord Jim), although grammatically redundant, had a higher semantic information content. The syntactic English was not only extremely high with regard to semantic information content but its grammatical redundancy was the same as the first order AE.

Treisman argues that the different regression slopes of DE (parts of speech) and the other DE measures imply that syntactic and semantic cues make independent contributions to word guessing. However, the difference in slopes may simply be due to the different ceiling of information content between the various measures which was mentioned above.

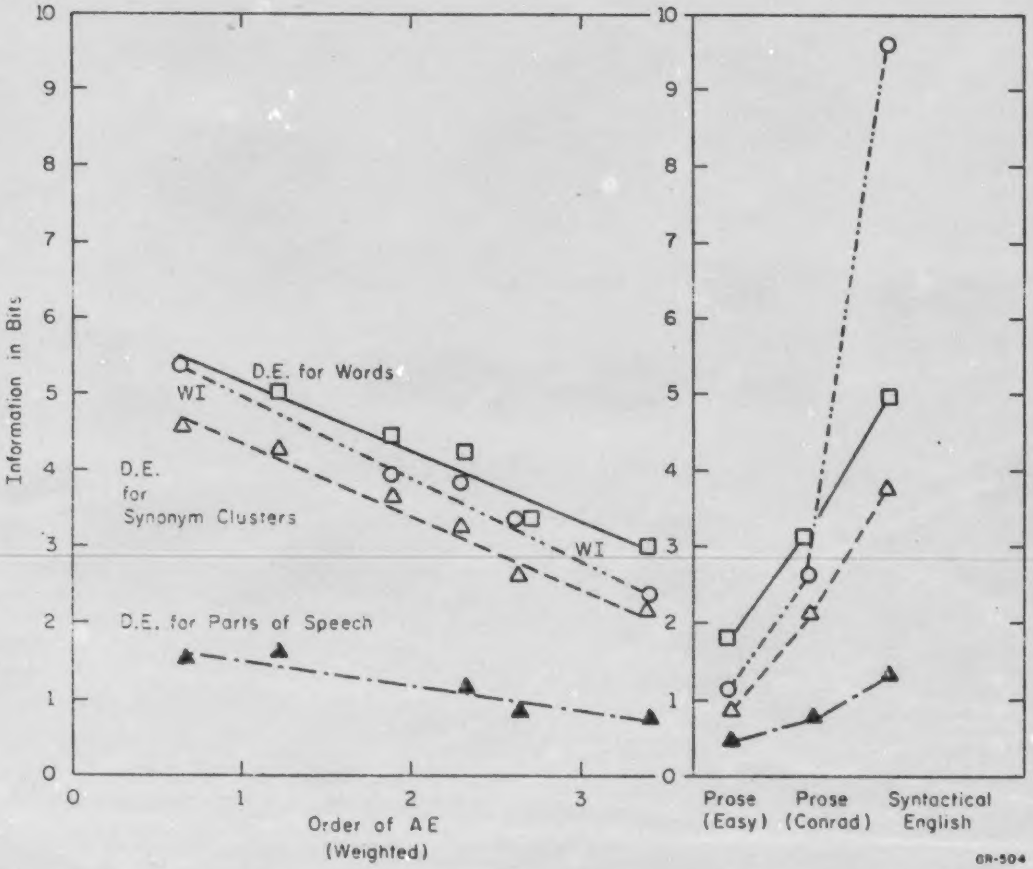


Figure 1. Treisman's (1965) results: D.E. in bits for various measures plotted weighted order of AE, and other passages.

Moreover, the results from the syntactical English suggest a rather intimate relation between meaning and structure. It seems that the overall conceptual coherence of a passage is of great importance, and the inability of Ss to guess the parts of speech correctly in the syntactical English condition implies that meaningful associations come first, and are prior to the structure, at least in a prediction task of this kind.

One factor for which Treisman did not control is word frequency. It was remarked above that the Dale-Chall index, which is based on word frequency, is correlated with both learnability and predictability. It is possible that the differences in predictability between the passages from Lord Jim and the easy prose sample, may simply be due to differences in word frequency between the two passages. Aborn & Rubenstein (1958) have shown that when Ss are allowed several attempts to replace a deleted word, they begin with common words, and move on to more infrequent words. It is quite possible that sequential semantic constraints and word frequency are both operative in Cloze tasks, but no one (to my knowledge) has as yet separated these two variables. Some evidence that semantic constraint per se is important is provided by Shepard (1963) who gave Ss a fixed period of time to generate as many replacements as possible for a deleted word in sentences of varying length. He found that increasing the amount of context sharply reduced the rate of generation of replacement words. He also confirmed Aborn, Rubenstein & Sterling's (1959) finding that contextual effects were asymptotic after the length of the bilateral context was increased beyond ten words.

The studies reviewed in this section show that short term sequential dependencies (semantic and structural), overall context, word frequency and form class are all important in helping to guess words that have been deleted from a text. Predictability is also related to learning. Rubenstein & Aborn (1958), whose experiment was discussed above, found a moderately high correlation between predictability and learning.

Slamecka (1946b), using the successive guessing technique, developed six 20 word sentences varying in predictability. An example of high predictability is: "The young child must learn that two and two are four, beyond the shadow of a doubt"; and an example of low predictability is "With many games, good balance or speed on a return swing produced what teams generally judge as being proper play." Slamecka found that predictability (or contextual constraint) exerted a strong influence on sentence acquisition, but a rather lesser influence on sentence retention, where remembering was assessed by re-learning after a six day retention interval. In general, it seems that predictability, whether assessed by the Cloze procedure or successive guessing, is an important variable in learning and remembering word strings.

Although the evidence which has been presented indicates that predictability depends on structural as well as semantic factors, the only structural factor as yet identified is form class. Sequential constraints have been referred to, but they have not been further described. Clearly, the operators which determine sequential structure are grammatical rules, and indeed there are studies which show that grammatical structure _per se_ facilitates learning.

Epstein (1961) required Ss to learn 6 different kinds of sentences, examples and descriptions of which are shown in Table 2. From the number of trials necessary to attain criterion, it can be seen that the addition of properly placed grammatical tags to a string of nonsense syllables significantly facilitates learning. From a second study (Epstein, 1962) he concluded that the facilitative effects of grammatical tags depend not on the establishment of any sequential associations, but on the fact that the tags allowed the word strings to be treated as units. In a related experiment, Forster (1966b) also found that grammatical tags assisted learning, and that the assistance was not mediated by merely providing cues to serial order.

An aspect of sequential constraint which has not been explicitly discussed is inter-item associations, although the semantic component of predictability must obviously be related to the associational relationships of the individual words. Although the operation of associations is constrained by context (Howes & Osgood, 1954), the presence of pre-established inter-item associations within sentences do contribute to learning (Johnson, 1968; Rosenberg, 1966, 1967a, 1967b).

Predictability is a response measure. It is a subjectively defined function of the grammatical structure of the sentence, and the characteristics of the words (absolute and contextual frequency of occurrence) that comprise the sentence. When we describe predictability, we are not describing an inherent property of a sentence that is independent of the outcome of the processing, or interpreting of that sentence by a person. In a sense, the predictability of a sentence is an intervening measure,

TABLE 2

SENTENCES USED by EPSTEIN (1961): examples

	Trials to criterion
I. nonsense syllables with grammatical tags plus function words	
"A vapy koobs desaked the citar molently um glox nerfs".	5.77
II. same as I, less the grammatical tags	
"a vap koob desak the citar molent um glox nerf".	7.56
III. Items of I in a random order.	
"Koobs vapy the um citar nerfs a molently."	8.15
IV. Same as I, with the grammatical tags channged	
"A vapy koobed desaks the citar molents um glox nerfly."	6.90
V. Anomalous sentences	
"Cruel tables sang falling circles to empty bitter pencils."	3.50
VI. Randomized words	
"Sang tables bitter empty cruel to circles pencils falling."	5.94

one that follows its logically defined construction, and precedes its learning by a particular person.

We have noted that a sentence's structure is described by its grammar, and that grammatical factors are important in learning and retention. The next step is to explain the grammatical descriptions of sentences a little further.

Grammatical models of sentence structure.

There are three kinds of generative grammars which are capable of producing English sentences (Chomsky, 1957). They are finite-state grammar (FSG), phrase-structure grammar (PSG), and transformational grammar (TG).

Finite-state grammar (FSG). A FSG is a Markov source, in that the process of formulating sentences is viewed as a series of selections of words, each selection being determined by the antecedent words. This model generates sentences in a left-to-right fashion, each word depending only on the preceding words. It is apparent that the grammar pre-supposed by Miller & Selfridge (1950) is finite state, as the selection of any particular word depends on the preceding words. Although the matter has not been made explicit, predictability studies which are simply concerned with the preceding amount of context are based on some form of FSG.

One attraction of a FSG for psychologist is that it is quite closely related to association theory. For example, Hull (1930), proposed a chaining theory to account for sequential activity in which each response produced a stimulus for the next response. Essentially, his model may be represented by a simple Markov transition matrix. Staats & Staats (1966)

S-R model of language production also may be represented by two sets of Markov processes, one set dealing with the sequential probability of grammatical form class, the other, embedded in the first, dealing with the words within these form classes.

However, FSG has some drawbacks. It is not *in principle* adequate as a grammar of English (Bever, 1968; Chomsky, 1957, 1959; Fodor & Bever, 1968; Miller & Chomsky, 1963; Miller, Galanter & Pribram, 1960). It is counter-intuitive, in that most people are aware of the endings of their sentences when they are still halfway through speaking them. Aborn *et al*'s (1959) and Shepard's (1963) finding of bilateral influence on predictability would not be expected by an FSG. Epstein's (1962) and Forster's(1966b) findings that grammatical tags have an effect over and above the facilitation of sequential associations also suggest that a FSG is inadequate. Neither do FSG's provide a way in which chunks may be easily inferred (order of AE introduces artificial fractures into the word strings that are not normally present in the natural language), and we have already cited evidence which shows the importance of chunking.

A consequence of the necessary inadequacy of FSG (it has been proved by Chomsky that it cannot generate all and only all the grammatical sentences in English) means also that association theories are intrinsically incapable of providing a complete account of language behavior. That is not to say that associations and statistical probabilities of responses are not important; they demonstrably are. However, they are not sufficient, and so we must look further for an adequate structual description of sentences.

. Phrase-structure grammar (PSG). The claim that a set of elements is structural means that a change in one element creates systematic changes in others, even if these others are remote from the first. This claim is certainly true of sentences; their elements are not independent, and the problem of measuring sentence structure has arisen from the question of how this inter-dependence should be assessed. A further complication in the case of sentences is that not all inter-relationships are equally strong; the structure of sentences is hierarchical.

A descriptive PSG takes structural units (properly morphemes, although parts of speech will serve in this instance) and expresses the relationship between them either by nested labelled bracketing or by a tree diagram (which is also called a phrase-marker or P-marker). Both descriptions are illustrated in Figure 2. They are actually a special case of the set of possible kinds of P-markers, called a binary P-marker, which is obtained by immediate constituent (IC) analysis.

The basic procedure that is used in an IC analysis is to take a sentence, and divide it into two units. This division usually divides the sentence into the subject and the predicate. At the next level down these two constructions (the subject and the predicate) are themselves divided into their major constituents, and so the process continues until the whole binary tree is complete. The process is illustrated in Figures 2 and 3.

One virtue of PSG is that it can distinguish rather easily between many sentences which have exactly the same word sequence (and are thus identical from an associationistic view), but are actually

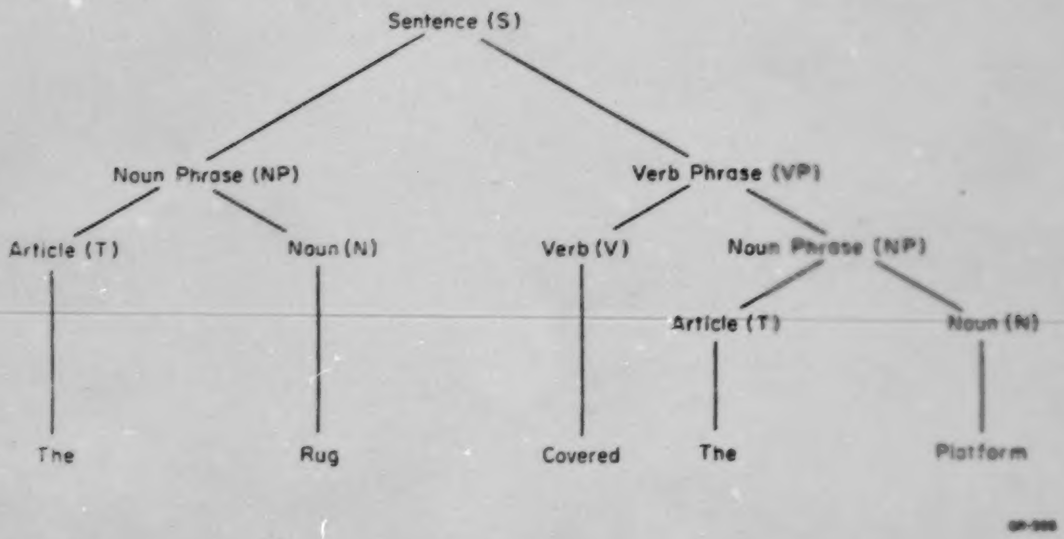

Sentence (S)
Noun Phrase (NP)
Verb Phrase (VP)
Article (T)
Noun (N)
Verb (V)
Noun Phrase (NP)
Article (T)
Noun (N)
The
Rug
Covered
The
Platform

Figure 2. An example of constituent structure analysis.

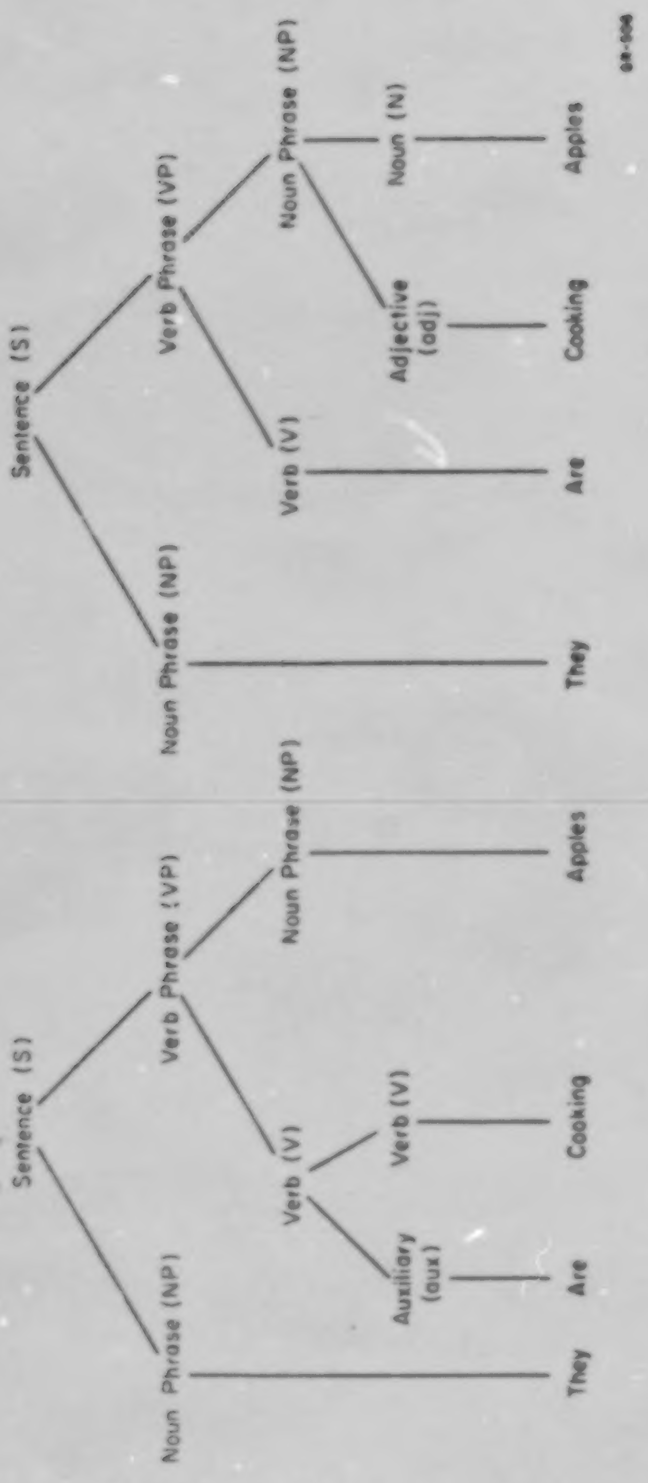

Sentence (S)
Noun Phrase (NP)
Verb Phrase (VP)
Verb (V)
Noun Phrase (NP)
Auxiliary (aux)
Verb (V)
They
Are
Cooking
Apples
Sentence (S)
Noun Phrase (NP)
Verb Phrase (VP)
Verb (V)
Noun Phrase (NP)
Adjective (adj)
Noun (N)
They
Are
Cooking
Apples

Figure 3. An example of how an ambiguous sentence may be resolved by means of constituent structure analysis.

ambiguous. For example, consider the sentence:

(1) They are cooking apples.

The word cooking may be either a present participle or an adjective. The alternative meanings can be readily expressed by bracketing the words into successively larger units, or by means of a tree diagram. The alternative meanings are shown in sentences 2 and 3 below and also in Figure 3.

(2) ((They) ((are cooking) (apples))).

(3) ((They) ((are) (cooking apples))).

A way in which P-markers could be used in a model of sentence generation has been devised by Yngve (1960). Consider the sentence:

(4) The old man has very weak knees.

Yngve assumes that initially this sentence might be represented by some symbol S, leaving the nature of S (image, expression, concatenation of neurals events, pride of $r_m s$) unspecified. He then generates the sentence by means of so-called binary rewrite rules, which start at the topmost node (S) of a sentence, and proceed to work down through successive nodes until the sentence is generated from left to right. The process, which is really the reverse of IC analysis, is illustrated in sentence 4 in Table 3. The tree diagram which is generated by the continued application of these rewrite rules is shown in Figure 4.

Yngve was interested in problems of mechanical translation, but Martin & Roberts (1966) have suggested that his model can be readily be coordinated to a psychological model of the listener and the speaker. The occurrence of any one word in a sentence (whether uttered or heard)

TABLE 3

Rewrite operations necessary to generate The old man has very weak knees. The rewrite (or decoding) operations themselves are preceded by an asterisk; other operations merely translate word classes into words. To the right of each operation the specific nodes being held in memory are shown.

Operation		Nodes stored in memory
* S	NP_1 + VP	------
* NP_1	T + NP_2	VP
T	The	NP_2, VP
* NP_2	AdjP + N_1	VP
Adj	old	N_1, VP
N_1	man	VP
* VP	V + NP_3	------
V	has	NP_3
* NP_3	AdjP + N_2	------
* AdjP	Adv + Adj_2	N_2
Adv	very	Adj_2, N_2
Adj_2	weak	N_2
N_2	knees	------

S = sentence, NP = noun phrase; VP = verb phrase; T = article; Adj = adjective, N = noun, V = verb, AdjP = adjective phrase; Adv = adverb

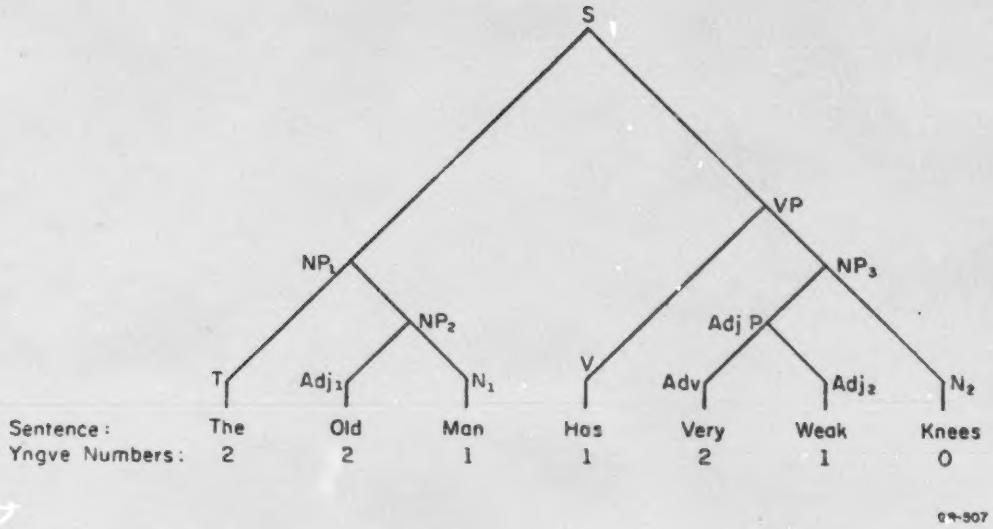

S
VP
NP1
NP2
NP3
Adj P
T
Adj1
N1
V
Adv
Adj2
N2
Sentence: The Old Man Has Very Weak Knees
Yngve Numbers: 2 2 1 1 2 1 0
84-507

Figure 4. The tree diagram or P-marker of the illustrative sentence. The old man has very weak knees. The Yngle numbers associated with each word are also shown.

carries implications for any following words. For example, in sentence 4, the first word, the, implies a noun phrase and a predicate. The second word, old, implies the completion of the noun phrase and a predicate. Similarly, the other words imply succeeding words of certain particular classes.

Now it is clear that some words have more implications for the remaining words than others; for example, the first is always pregnant with sentential possibility, while the final word is always barren. It may be conveniently assumed that at the time any particular terminal word is written out (or received) there exists a set of expectations about what classes of words are to come, these expectations (or in the case of the speaker, intentions) being carried in short-term memory (STM). The finite capacity of STM is well established, and so it comes as no surprise that the number of nodes which may be implied at any one time are quite limited. This suggests that the average number of nodes in STM during the course of a sentence might provide an index of perhaps comprehension or learning difficulty. Equally well, the number of nodes in STM for a particular word may be said to reflect the depth of structural embeddness of that word.

In the case of sentence 4, the number of nodes in STM for each terminal word are shown in Table 3. The number of nodes in memory for a word may also be obtained by counting the number of branches leading to that word in the tree diagram (Yngve, 1960). Martin & Roberts (1966) have suggested that such numbers be called Yngve numbers, and we will follow their suggestion. Computations of Yngve numbers are shown in

Figures 4 and 5.

Some basic properties of Yngve numbers are worth noting. First, they carry only structural information; such factors as word-class uncertainty are not taken into account. This is because Yngve numbers are assigned to word classes only on the basis of structural relations between word classes. Second, an Yngve number must always be a non-negative integer: it is zero only for the final position in the sentence. If Y_i is the Yngve number of the i^{th} position, then Y_i-1 cannot be larger than Y_i+1; and if there are n words in a sentence, the largest Yngve number cannot be larger than n-1. Given these properties, the mean depth of a sentence is bounded as follows (where mean depth is equal to the average of all Yngve numbers of the separate words, and which may be written as Y):

$$(n-1)/n \quad \overline{Y} \quad (n-1)/2$$

Yngve's index is not the only possible measure of sentence complexity that may be derived from PSG P-markers, nor is it entirely adequate. Structural embeddedness is defined in such a way that left-branching, self-embedding and multiple branching all contribute equally to depth. However, Miller (1962) and Miller & Isard (1964) have shown that a multiply self-embedded sentence is particularly difficult to understand. But this evidence does not, of course, invalidate the application of Yngve measure to other sentences.

Transformational Grammar (TG). Whatever the usefulness of PSG to psychological theory, it is now generally agreed by linguists that it is in principle insufficient to provide an adequate generative model

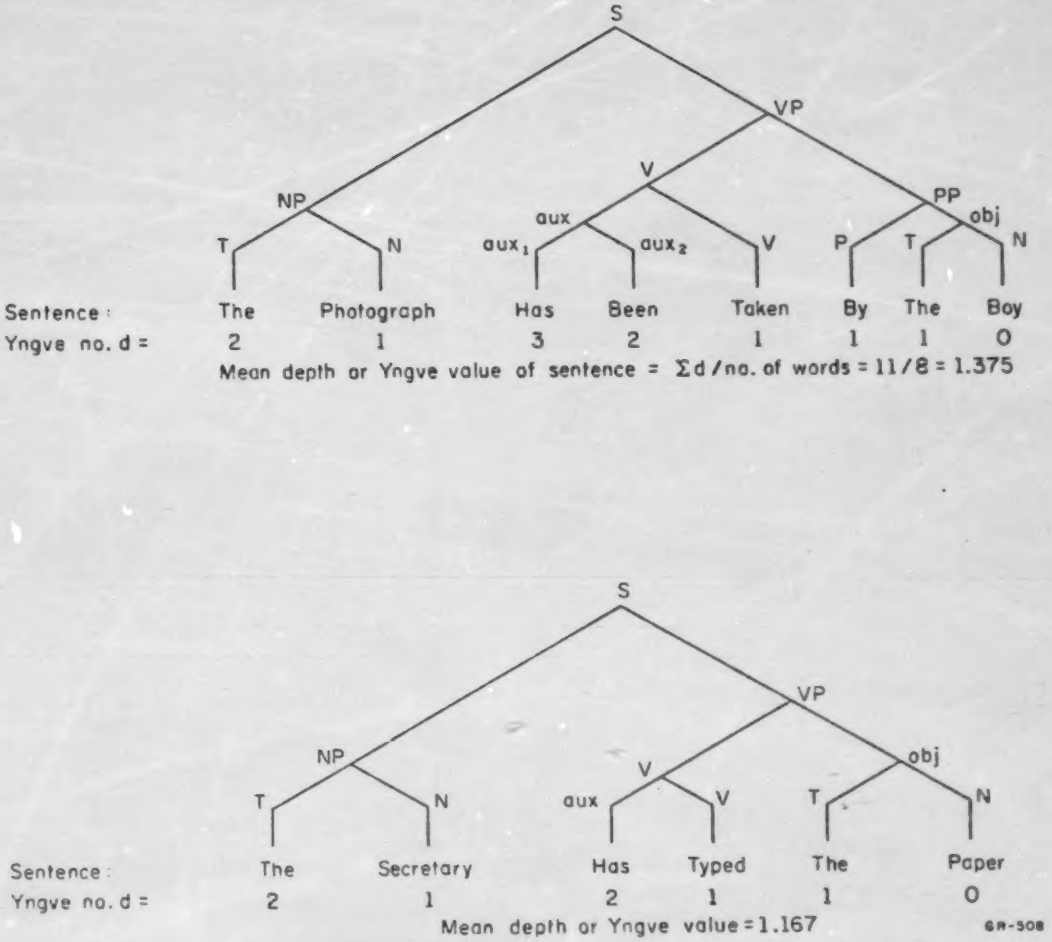

S
VP
V
NP
PP
obj
aux
T
N
aux₁
aux₂
V
P
T
N
Sentence:
The
Photograph
Has
Been
Taken
By
The
Boy
Yngve no. d =
2
1
3
2
1
1
1
0
Mean depth or Yngve value of sentence = Σd / no. of words = 11/8 = 1.375
S
VP
NP
obj
V
T
N
aux
V
T
N
Sentence:
The
Secretary
Has
Typed
The
Paper
Yngve no. d =
2
1
2
1
1
0
Mean depth or Yngve value = 1.167
GA-508

Figure 5. An illustration of the way Yngve values are calculated. The Yngve number of a particular word is its depth of embedding in the sentence, and is the sum of the left branches of the tree at each node, when proceeding from the top of the tree down to the word in question.

of language. Certain kinds of sentences cannot be properly handled by a PSG. For example, consider:

(5) John is easy to please

(6) John is eager to please

and

(7) John hit the ball

(8) The ball was hit by John

The first pair have the same surface structure (i.e. their trees are identical), but a different meaning, while the second pair have different phrase structures, but the same meaning. It has been proposed (Chomsky, 1957, 1965, 1966; Miller, 1962), that each actual or surface sentence is derived from some underlying base sentence, and that the sentence which is written or spoken is obtained by transforming the base sentence. Thus in the examples it is argues that the base sentences in 5 and 6 are different, but that in 7 and 8 they are the same.

A satisfactory syntactic description should determine uniquely the semantic interpretation of a sentence (i.e. what the sentence means). The deep structure (which contains the base sentence) determines the semantic interpretation, and the surface structure determines the phonetic form. The grammar, according to transformational grammarians, then consists of a syntactic component, which generates syntactic descriptions of deep and surface structures, a semantic component, which assigns a semantic interpretation of the deep structure, and a phonological component, which assigns a phonetic interpretation to a surface structure.

TG has undergone one major revision since its first formulation. Since the first version has had a great deal of impact in psycholinguistics, it is fitting to describe it. Consider the sentence (from Chomsky, 1966)

(9) I expected the man who quit work to be fired.

It is assumed that the sentence is synthesized in the manner shown in Figure 6. The base component generates three base P-markers, properly represented by symbols, here shown as sentences. These sentences are, I expected it (B1), Someone fired the man (B2), The man quit work (B3). A series of transformation are then performed on these base sentences (actually, symbol strings). First, a relative transformation (T_{rel}) changes B3 to Wh-(the man) quit work. Second, an embedding transformation (T_{emb}) combines B2 and B3, and deletes the man, to get Someone fired the man who quit work. Third, a passive transformation (T_{pass}) converts B2 and B3 to the man who quit work was fired by someone. Fourth, another transformation (T_{del}) deletes by someone. Fifth, T_{emb} is applied to B1 and B2 and B3, to produce I expected the man who quit work was fired. Finally, a tense transformation changes the final was to to be, to arrive at sentence 9.

It may be noted that some transformations are obligatory if the sentence is to be grammatical, like T_{rel}, T_{emb}, and T_{del}, whereas others are optional, like T_{pass}, and T_{te}. A so-called kernel sentence is obtained when only obligatory transformations are applied.

This version of TG has been roundly criticized (Chomsky, 1965), with the result that the new model introduces all meaning bearing elements into

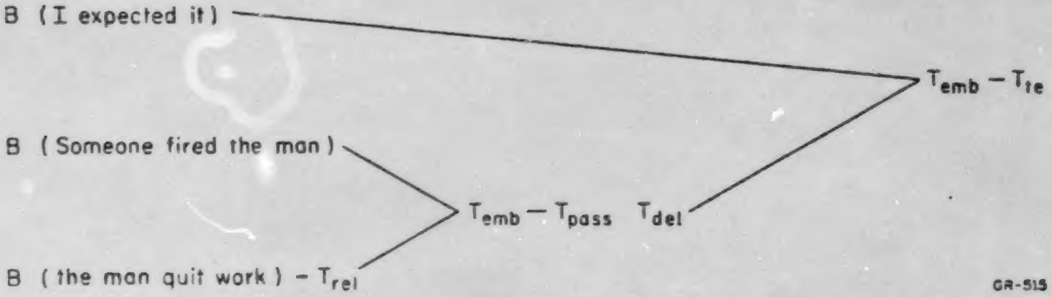

B (I expected it)
B (Someone fired the man)
B (the man quit work) - T_rel
T_emb — T_pass T_del
T_emb — T_te
GR-515

Figure 6. (Following Chomsky, 1966). A T-marker showing the transformational derivation of I expected the man who quit work to be fired, according to the first version of TG. (see text for explanation).

the base, with the task of the transformations being to inter-relate the sentences in the generalized base P-marker, (in the example) in Figure 6, a generalized P-marker would embrace B1,B2, and B3. The procedure is illustrated in Figure 7.

Transformations operate first on the most deeply embedded structure, and work their way up. First T_{rel} is applied to the bottom phrase, marked A. Then, turning to phrase B, the relative clause and the following noun, man, are inverted. Next, phrase C is integrated, and the dummy subject, someone, is deleted. Note that the passive, a meaning bearing element, is incorporated into the P-marker. Finally, the remainder of the sentence is integrated, and the tense changed. The result is sentence 9. The critical change between this and the previous version is that all the semantic information is contained in the generalized base P-marker.

The foregoing is relatively non-controversial. The same is not true about the questions of how the base P-marker is generated, and how it is given a semantic interpretation (Chomsky, 1966; Weinreich, 1966). It is suggested that the base (which produces the base P-marker) consists of two parts, a categorial component, and a dictionary or lexicon. The categorial component consists of rewriting rules which are defined on either category symbols (nouns, adjectives, verbs, etc.), complex symbols (category symbols and a matrix of semantic features, e.g. noun: + animate, + count), and dummy symbols.

The categorial component generates a pre-terminal string of symbols. The lexical items are then inserted according to two criteria. The first is strict sub-categorization, which specified the syntactical restraints on an item, e.g. noun, transitive verb, etc.,

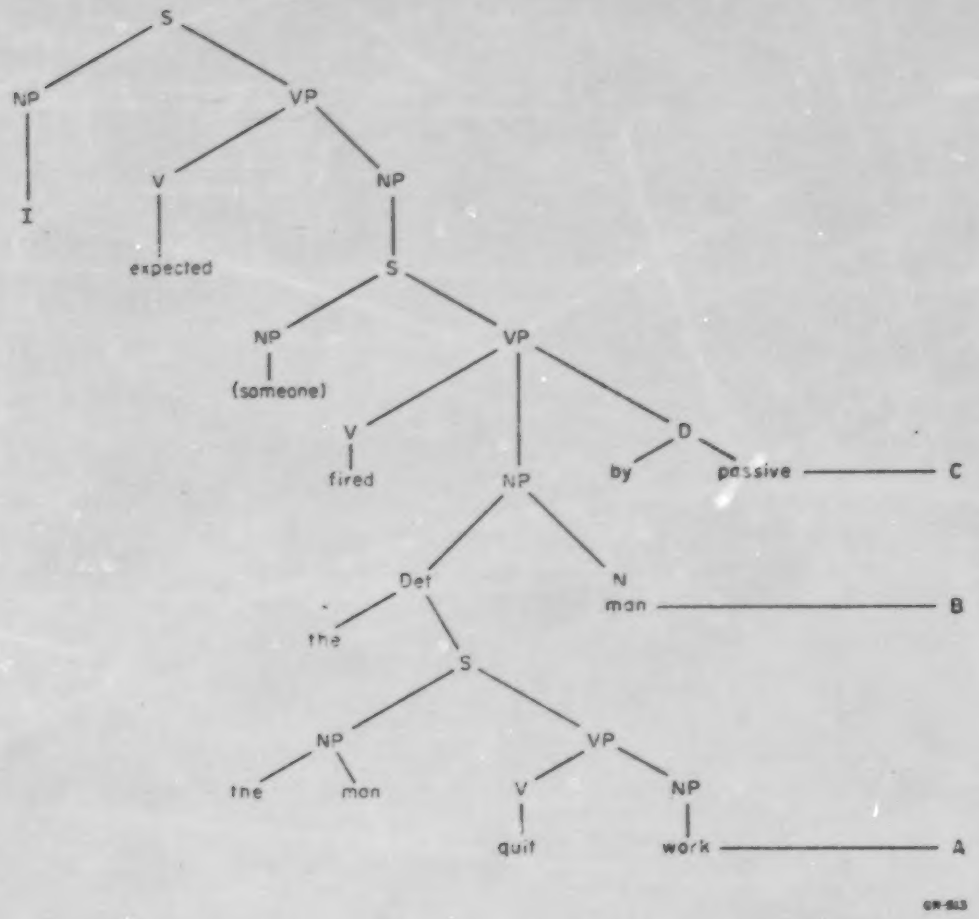

S
NP
VP
I
V
NP
expected
S
NP
VP
(someone)
V
NP
D
fired
by
passive
C
Det
N
the
man
B
S
NP
VP
the
man
V
NP
quit
work
A

Figure 7. (Following Chomsky, 1966). A generalized base P-marker of *I expected the man who quit to be fired* (see text for explanation) and the second, selectional congruence, which specifies the semantic restraints on an item, e.g., the phrase, pregnant stone violates semantic constraints. The flow chart of such a process is shown in Figure 8. The TG model is much more ambitious than PSG or FSG, and of course, in a sense it includes these latter two descriptions. We have already shown that although FSG may be inadequate as a grammar, when realized as association theory, or AE, it can be quite useful in explicating psychological processes. Similarly, the surface structure may provide indices of learning difficulty (Martin & Roberts, 1966), and it is certainly important in sentence perception (Miller, 1962; Neisser, 1967, Pp. 259 ff). The question now is: does TG have any value as a model of psychological functioning?

If the meaning of a sentence (that part of it which is important for understanding or comprehension) is found in the abstract underlying base structures, then speed of understanding should be related to the number of transformations which intervene between the base and surface P-markers. This prediction has been widely tested and with some qualification, has usually been confirmed (Clifton & Odom, 1966; Clifton, Kurcz & Jenkins, 1965; Gough, 1965; Mehler, 1963; Miller, 1962; Miller & McKean, 1964; Slobin, 1966; Ervin-Tripp & Slobin, 1966). In a subsequent section concerned with theories of memory, we will discuss this and other evidence in more detail.

Theories of memory and some related evidence.

Our discussion began with questions about the ways in which sentences might be described in order that the retention of sentences

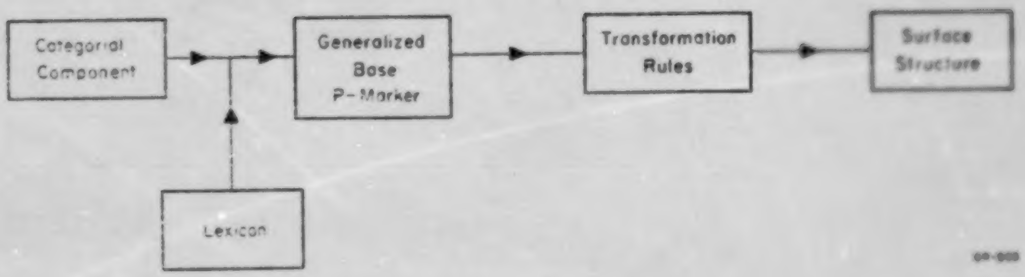

Categorial Component
Generalized Base P-Marker
Transformation Rules
Surface Structure
Lexicon

Figure 8 Flow chart of TG sentence generating process

could be studied. We have seen that meaning density, chunkability, predictability (syntactic and semantic) and structure are all important. It is now time to consider some theoretical conceptions of memory, some of which grow out of the previous discussion, and related evidence.

Traditionally, memory has been regarded as a complex network of associations, the elements and their bonds being conceived according to classical association theory (Bahrick, 1966; Humphrey, 1951, chap. 1; Wales & Marshall, 1966). It seems indisputable that memory is associative, with discrete elements (specific memories) being stored, but it is equally beyond doubt that memory has dynamic and structural characteristics as well (Bartlett, 1932; Neisser, 1967). Indeed, for some, memory _is_ organization (Mandler, 1967, 1968). But there is a great deal of evidence which shows that associative connections are important (Adams, 1967; Cofer, 1966, 1967). The truth of the matter is probably that associative relationships are both organized and structured, and it is the extent of the role of the organizing or structural factors which is interesting.

In the first theory that we will consider it is argued that memorizing involves abstracting information from the input, storing the abstracted (and re-organized information), and then reconstructing the input at recall. Bartlett (1932) and Katona (1940) were among the originators of abstractive theory, but the most recent statement of the general idea comes from Gomulicki (1956). Bartlett found that in the recall of prose passages many details were omitted, others changed, and still others inverted. He concluded that the past operates as an organized mass rather than a group of elements, each of which retains its specific character. However, as Katona realized, it is possible to remember detail, and indeed fixed trace, or association theories have enjoyed their greatest successes in the field of learning. Gomulicki argued that what is needed is a flexible trace theory. Using prose passages of varying length (13-95 words) and immediate free recall, he found that Ss omitted material in a highly selective fashion. Unimportant words, phrases, and even sentences were deleted, indicating that an abstraction process (called mnemic abstraction by Gomulicki) was operating immediately on incoming material. This same process had been noted by Henderson (1903), and Lewis (1933), but not by Bartlett, who, in observing repeated recalls, was more strongly impressed by the vulnerability to distortion or forgetting of every aspect of the input.

Gomulicki's findings imply that an S is able to decide immediately which parts of a sentence are important, and abstract accordingly. In fact, he found that verbs were best remembered, followed by agents (the logical subjects), followed by the recipients of an action. Items which retarded the narrative of a passage, such as purely descriptive sentences, were remembered least well. These results led Gomulicki to conclude that what is remembered is an action-agent-effect unit, which in Gestalt terms comprises the figure of a passage, the remaining words constituting the ground.

Evidence supporting Gomulicki's view comes from Mandler & Mandler (1963). They presented sentences as serial lists, a word at a time, using an anticipation procedure. They found that the normal U-shaped serial learning curve did not appear; whether or not a word was learned quickly rather depended on its contribution to the core meaning of a sentence. The core meaning was not always organized into syntactically valid units either; any meaningful combinations of words could serve, but usually the core meaning was carried by an action-agent-effect unit. A typical result is shown in Figure 9. Pompi & Lackman (1967) propose a theory which is similar to Bartlett's. They write:

> "The point of view taken here is that a fundamental characteristic, and perhaps the central mystery, of meaningfully connected discourse, is that is generates surrogate structures which are not absolutely dependent upon the verbal material. The structure of the discourse is critical: a surrogate system (some combination of theme, image, schema, abstract, or summary) depends for its appearance upon, among other things, the word order of the meaningful material....
> Implicit in this statement is the view that an agrammatical list of words may generate an "essential idea" or surrogate process if associative relationships of sufficient magnitude exist between the word". (Pompi & Lachman, 1967, Pp. 143-144).

They carried out two experiments, testing the recall and recognition of 75 word passages of connected discourse. They found that the data were consistent with the view that S̲s store surrogate structures; in other words, S̲s recode material into semantic chunks, an argument which is very close to that advanced by Henderson (1903) about the importance of IGs or meaning units. In Pompi & Lachman's formulation though, the surrogate structure may use different lexical items from the original.

A second theory of memory comes from TG. In essence, the argument is that what is stored is the deep structure of a sentence, and so

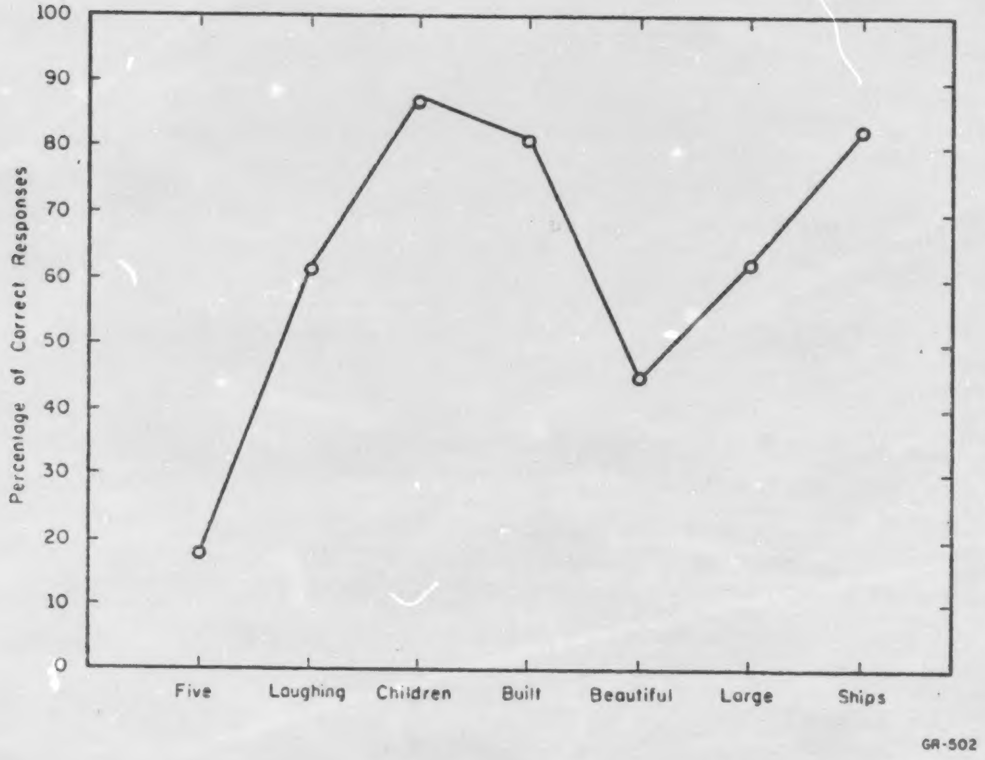


Figure 9. An illustrative example from Mandler and Mandler (1963). (See text for explanation).

learning involves recovering the deep structure from the surface structure, and recall requires the reverse process. It follows that, insofar as sentential complexity is a function of syntactic variables, the complexity of a sentence is measured by the number of grammatical transformations intervening between the deep and surface structures. A plausible addition to this hypothesis is that the meaning of a sentence and its syntactic characteristics are in some wise stored separately (Miller, 1962).

Marks & Miller (1964) gave Ss a set of 20 sentences in an immediate free recall task. They used four kinds of word strings, grammatical and meaningful sentences, grammatical but semantically anomalous sentences, anagram strings, and random word strings. Examples of each kind are shown in Table 4. Semantically anomalous sentences were created by replacing words in a particular sentence with other words of the same form class from elsewhere in the list of 20 sentences. This procedure resulted in a violation of selectional restrictions. Anagram strings were created by rearranging the word order of grammatical sentences, resulting in a violation of rules of strict sub-categorization (as well, in some cases, as selectional restrictions). The random word strings were obtained by randomly choosing individual words from various of the 20 sentences.

The dependent variables (obtained from the free recall test) were total number of words correct regardless of position, number of strings completely correct, and the number of words which were correct and properly positioned. Overall, normal sentences had many fewer errors than the other word strings, anonalous sentences and anagram strings had

TABLE 4

Examples of the types of sentences by

Marks & Miller (1964).

SENTENCE TYPE	EXAMPLE
Grammatical:	Gallant gentlemen saved distressed damsels
Anomalous:	Gallant detergents fight accurate fumes
Anagram String:	Distressed gallant saved damsels gentlemen
Random Words:	Accurate gallant fight fumes detergent

approximately the same number of errors, and the greatest number of errors were due to the random word strings.

Marks & Miller also looked into different kinds of errors; in particular, those of inversion, bound morpheme location and intrusions. A summary of the distribution of error types is shown in Table 5. They argued that intrusions can be considered as semantic errors, related to decisions as to which words may combine in a sentence, and thus they occur most frequently in anomalous sentences and word lists, where semantic rules are violated. Bound morpheme errors and inversions can be considered as syntactic errors; the first related to grammatical tags, and the second to word order. It may be seen from looking at Table 5, that these occur most frequently in anagram strings and word lists, where syntactic rules are violated. The association of semantic errors with semantic violations, and syntactic errors with syntactic violations, led Marks & Miller to conclude that meaning and syntax are coded separately.

Mehler (1963) compared immediate recall of differently transformed sentences. He used active or kernel (K), passive (P), negative (N), and interrogatory (Q) transformations, as well as the following combinations of these, PN, NQ, PQ, and PNQ. Recall was aided by a prompt word, either the subject or object of the sentence. He found that kernel sentences were learned more quickly and that errors that did occur were usually ones of simplification. He concluded that the transformation applied to a sentence was coded separately from the sentence itself.

Savin & Perchonock (1965) used a task based on Archimedes'

TABLE 5

Summary of Marks & Miller's (1964) results for error types.

(Shown are the mean proportion of errors over all trials per S)

SENTENCE TYPE	ERROR TYPE		
	Inversions	Bound Morpheme Errors	Intrusions
Grammatical	.02	.13	.06
Anomalous	.07	.27	1.00
Anagram strings	1.94	2.63	.17
Random words	.30	1.44	.94

Principle in which Ss were presented with a sentence plus eight unrelated words, and then immediately had to recall both the sentence and the words. The number of words that could be recalled provided an index of the amount of space in short term memory that was occupied by the sentence. They found that the number of words recalled depended on the number of transformations contained in the sentence, and so concluded that a kernel sentence was stored separately from any transformation that was applied to it.

Clifton, Kurcz, & Jenkins (1965) presented Ss with a test list of 96 sentences, 48 of which had been seen previously, and the Ss had to press a key when they detected a sentence that they had seen before. The distractors (or new sentences) varied in that they were either the K, N, P, or PN form of the old sentences. They found some evidence that sentences which are grammatically similar are also similar psychologically. However, the relationship was not a simple one, in that the probability of falsely recognizing a new sentence as an old one was different for the different transformations, and the effects of PN were not a simple additive function of the effects of P and of N separately.

Further evidence about the relationships of the transformations imposed on a sentence to one another, is provided by Clifton & Odom (1966), and Slobin(1961). The first two authors found that in contradiction to Mehler (1963) and Miller (1962), a prism model rather than a cube model of the relationship between the various optional transformations obtains (See Figure 10). They re-analyzed Mehler's data and concluded that they, too, fit a prism better then a cube model.

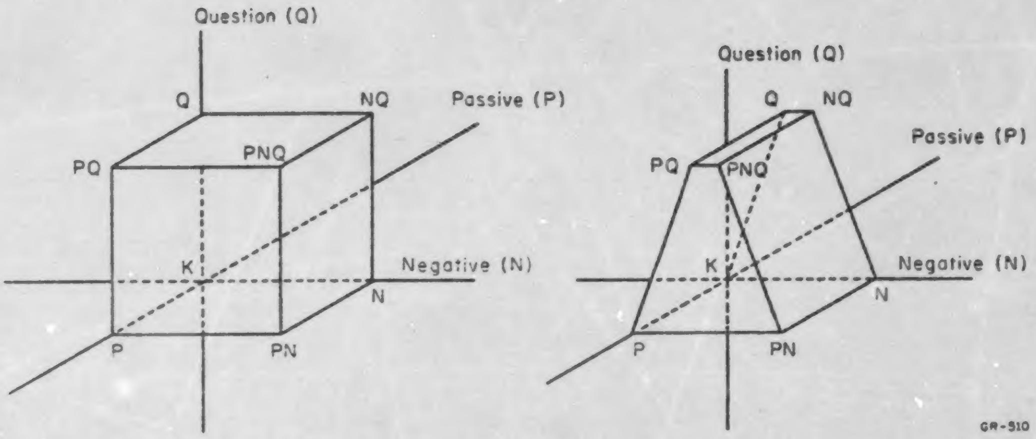

Question (Q)
Q
NQ
Passive (P)
PQ
PNQ
K
Negative (N)
N
P
PN
Question (Q)
Q
NQ
Passive (P)
PQ
PNQ
K
Negative (N)
N
P
PN
GR-510

Figure 10. (Following Clifton & Odom, 1966). A schematic representation of the cube and prism models, both of which purport to describe the relation between the various transformational operations.

Slobin (1966) found that syntactic and semantic factors interacted when sentences were verified with respect to pictures. He found that syntactically simple negatives took more time to verify than relatively more complex passives.

The evidence supports the view that P, N, and Q carry a semantic component, an opinion arrived at independently by linguists. These so-called transformations really belong in the base. The experiments which we have discussed grew out of the first version of TG. In general, their outcome supports the second version.

These results do not, of course, inform the hypothesis that the complexity of a sentence depends (in part) on the number of transformations used in its derivation. They merely indicate that some operations presumed to be transformational were actually not.

A modification of this view comes from Fodor & Garrett (1967). They suggest that the complexity of a sentence is a function not only of the transformational distance from its base structure to its surface structure, but also of the degree to which the arrangement of elements in the surface structure provides clues to the relations of the elements in the deep structure. To a certain extent this view, and the previous one (that only the number of transformation are important) yield similar predictions as in general, increasing the distance from base to surface structure tends to excise cues to the former. However, they assert that cues to deep structure, and not the number of transformations involved is the important factor. This view also assumes that in order to understand a sentence it is necessary to apprehend the deep structure.

If it is some representation of the deep structure that is stored in memory, then it is conceivable that optional transformations, which simply realize the surface structure, are simply discarded, once the sentence is understood. Since the studies cited have used operations that were (on the basis of current TG) misnamed as transformation rules, they do not test this notion. Although the number of transformational rules or cues to deep structure should affect the rate of acquisition of a sentence, once the deep structure is recovered, they should be rapidly forgotten. This view, is, of course, similar in some respects to Gomulicki's.

Savin (1966) compared memory for surface structures derived either by self-embedding or by right-branching. Using the Archimedean technique of Savin & Perchonock (1965) he found that both types took up the same amount of space in memory. However, he found that recall for the self-embedded sentences was poorer, even though their sense was preserved, which suggested to Savin that the sentences were processed and stored in terms of their deep structure.

Sachs (1967) auditorily presented prose passages to S̲s̲. Passages ranged in length from 27 to 180 syllables, and at the end of a passage a test sentence was presented which was either similar to, or the same as a sentence in the passage. The number of syllables between the first occurrence of a sentence and its test was either 0, 80, or 160. An example of the variation between sentences is shown below. One of these would be a test sentence.

Basic sentence - He sent a letter about it to Galileo, the great

Italian scientist.

Semantic change - Galileo, the great Italian scientist, sent him a letter about it.

Passive change - A letter about it was sent to Galileo, the great Italian scientist.

Formal change - He sent Galileo, the great Italian scientist, a letter about it.

Sachs found that judgments were uniformly excellent when the interpolation interval was zero, but that after 80 syllables of interpolated material the ability to recognize syntactic changes had fallen away sharply, while semantic changes were still readily detected.

Her findings support any view which argues that what is stored in memory is some abstraction from the input. Her results provide evidence for Gomulicki's position, and to a much lesser degree, for the TG position. Although the Formal change was not detected, neither was the Passive, which involved a change in the base.

Sachs (1967) took no account of context, yet clearly context is very helpful semantically while offering virtually no grammatical assistance. It is possible that her_Ss, although not able to remem- the precise sentence, were able to determine whether or not a distractor sentence was a possible candidate. All of the syntactic variations were consonant with the context, but the semantic ones were not.

Bregman & Strasberg (1967) employed subjective reports as one of their sources of data in an experiment studying the use of grammatical transformations. They found that subjectively at least, Ss did not learn the transformations through the use of images related to the words in the sentence. They suggest, therefore, a "semantic recoding" theory. It requires that a distinction be drawn between the transmission code and the semantic message. The grammatical structure is presumed to be part of the transmission code, but not part of the semantic message, which is in memory. It seems clear that singulary transformations form part of the transmission code, and the base constitutes the semantic message.

That grammatical cues are important, regardless of the semantic inter-relationships of words in a sentence, was shown by Rosenberg (1966). He compared the effect of grammatical and ungrammatical word order, and various degrees of free-association strength between the words, on the acquisition of sentences. Both factors were significant and independent.

All of these studies have used short retention intervals. In addition, they have not controlled for level of learning, so the main conclusion is that grammar makes a difference to sentence acquisition, and that what is stored is primarily words with perhaps some grammatical markers attached. Gomulicki's hypothesis remains as the most plausible description of the end result, even though its achievement probably depends on grammatical complexity of some kind.

A third type of theory of memory for sentences comes from PSG and emphasizes the importance of surface structure. In this theory it is generally asserted that what is stacked away in store are chunks (Miller, 1956), and that chunk boundaries are derived from the surface structure of a sentence. We may begin by mentioning some contrary evidence to TG i.e. the psychological importance of transformational complexity.

Salzinger & Eckerman (1967) used nonsense syllables in quasi-sentences, an effect which they achieved by the judicious addition of prefixes and suffixes. They found that the effect of grammatical structure was significant (as opposed to random ordering) but that different transformations had no different effect. Martin & Roberts (1966), who used Yngve numbers as indices of surface complexity, varied complexity and transformational structure independently and found no increase in difficulty as transformations were added, they did find a distinct and constituent effect of increased complexity. Martin & Roberts also re-analyzed Mehler's (1963) results and found that they could be explained solely in terms of differences in Yngve complexity between sentences. Mehler also confounded sentence length with the number of transformational operations and Martin & Roberts (1967) have shown that sentence length is an important variable.

A PSG description has also been used by Johnson (1965 a & b) as the basis of a model for encoding and generating sentences. There are a number of illustrations in everyday language behavior which suggest that sentences may be encoded in units larger than discrete words. For example, there are stress and intonation patterns that span several morphemes, and there are hesitations of varying length between words and groups of words. There is plenty of evidence that pauses indicate encoding decision points in speakers (although there are other causes) and usually pauses are separated by several words (Goldman-Eisler 1958a, 1961a, 1961b, Lounsbury, 1954; Maclay & Osgood, 1959, Suci; 1967 Tannenbaum, Willams, & Hillier, 1965). Evidence for the importance of phrases come from Fodor & Bever (1965), who found that phrases were seemingly perceived as units, in that stimuli arriving at the receptors during the course of a phrase, are perceived as arriving at the receptors during the course of a phrase, is perceived as arriving either before or afterwards. Johnson suggests that there is a correspondence between the word sequences described by a P-marker, and the functional units into which people encode sentences.

One way to detect the language units that Ss use is to study the process of response integration as they learn grammatical sentences. Johnson (1965) argues:

> "If subjects do handle language in large units, they should integrate the units before they put the units together to integrate the sentence. That is, during the course of learning, the subjects should learn to go from one word to the next word within a unit before they learn to go from the last word of one unit to the first word of the next unit. Therefore, if subjects are scored (SIC) for the probability of going from a right word to a wrong word (i.e. a transitional error) for each transition within a sentence, the probabilities should be lowest for transitions within units and greatest for between-unit transitions." (p. 48)

In a PA learning task, in which the stimuli were digits, and the responses were sentences, Johnson found a highly significant relationship between transitional error probabilities (TEPs) and the surface structure of the sentences. Two examples of the relationships are shown in Figure 11. It may be noted that the TE pattern reflects within-phrase as well as between-phrase structure. Indeed, the level of the constituent division in an IC analysis (illustrated in Figure 12 was highly correlated with the TEPs.

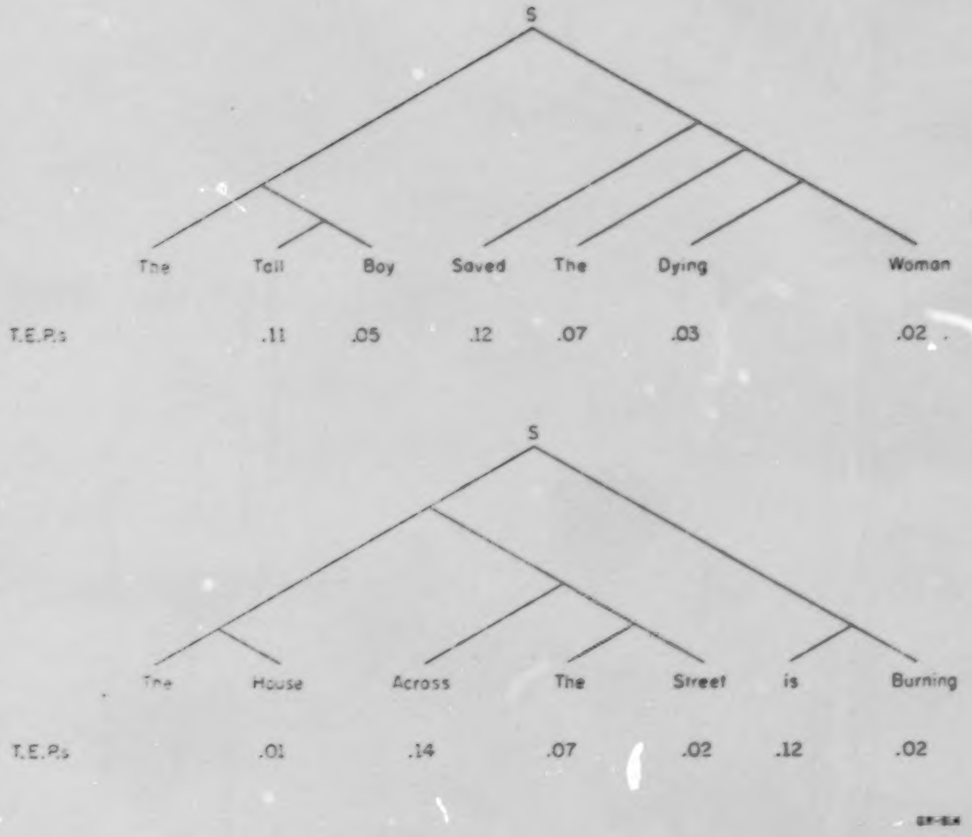

S
The
Tall
Boy
Saved
The
Dying
Woman
T.E.P.s
.11
.05
.12
.07
.03
.02
S
The
House
Across
The
Street
is
Burning
T.E.P.s
.01
.14
.07
.02
.12
.02

Figure 11. Data taken from Johnson (1965) illustrating the way in which transitional error porbabilities (T.E.P.s) are related to the surface structure of the sentence.

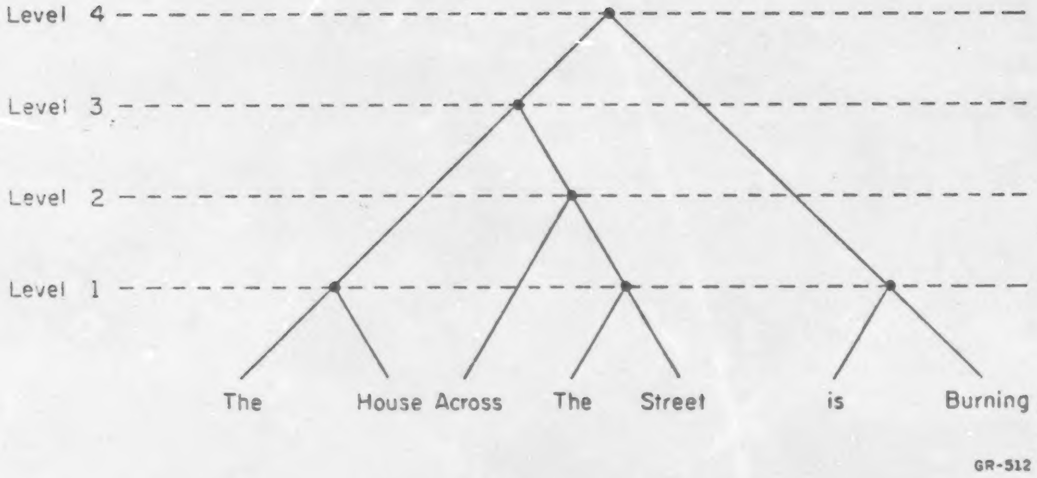

Level 4
Level 3
Level 2
Level 1
The
House
Across
The
Street
is
Burning
GR-512

Figure 12. An illustration of the way in which a tree diagram can represent different levels of integration of the words in a sentence.

Johnson (1965) reports a series of further experiments, which show his results could not be accounted for in terms of pre-existing word associations, although such associations certainly can have some effect, a qualification borne out by Rosenberg (1966).

However, it is doubtful whether a surface structure account is adequate. Blumenthal (1967) gave three presentations of a list of sentences to Ss, and then presented a prompt word for each sentence and asked the Ss to recall it. His sentences were either of the form "The child was warm by the stove" or "The child was warmed by the stove". The surface structure of these sentences is the same, but their underlying structure is quite different. In the first case, "by the stove" is an adverb of place, and in the second case it is the logical subject. When the prompt word was "stove" (or, in the other sentences, its grammatical equivalent), Blumenthal found that recall for the first kind of sentence was inferior to that for the second. In this case, the base structure provided correlates of psychological processes that did not exist in the surface structure. These results have been confirmed in a subsequent study by Blementhal & Boakes (1967).

Martin, Roberts & Collins (1968) presented sentences that were either active or passive, of either high or low Yngve complexity, and retention intervals of 0, 10, 20, or 40 sec. During the retention interval Ss counted backwards by 3s. They found forgetting curves that were similar in appearance to curves for single items like trigrams and word, although the error rate for sentences was much greater. Structural complexity (Yngve depth) was not related to the retention interval, but mean depth interacted with sentence kind. Among passive sentences less structural complexity allowed better recall; the reverse was true for active sentences. These results were not expected.

They also found that their results were correlated with differential word form class errors. Martin _et al_ argue that word classes are differentially processed into memory, and that the rules governing this selective processing are the rules of grammar. Subjects selectively focus on key elements of the input string, with grammatical structure acting as the functional stimulus which directs selection. Recall, then, would consist of generating a grammatical English sentence that incorporates specifically the key elements. This provides support for Gomulicki's position and little evidence for either of the other theories.

A methodological reflection on this experiment is fitting, Martin _et al_, in using sentences rather than single units, introduce a complication into the experimental paradigm which they used. Particular words early in the sentence have a longer time to be processed into memory than words at the end of the sentence. Crawford, Hunt & Peak (1966) have shown that retention of complex material improves over time: time to organize is important. Sentences whose finishes are predictable should be easier to remember than sentences which are not.

Summary. There are several views on how sentences and prose are stored in memory. Their basic differences may be most easily shown by a diagram.

From Figure 13 it can be seen that there are roughly two notions about how sentences are entered into memory one uses a decoding process and the other an abstracting process. The decoding view is held by Johnson (1965) and Miller(1962). Essentially it is that a sentence is resolved into its structural constituents in some fashion, either into a kernel sentence plus grammatical tags, or into some "image" from which the sentence can be later encoded (but see Osgood, 1968, Pp. 515). The difference between Johnson and Miller is that the former believes that all the information necessary to code a sentence is present (as required) in the surface structure, while the latter sees the outward and visible sentence as a starting point from which to begin the search for the underlying sentence. The abstraction view is held by Gomulicki (1956) and Martin, Roberts & Collins (1968). Their position is that the grammar provides a means of identifying the key words which provide the core meaning of the sentence, and it is the core meaning of the sentence, and it is the core meaning which is stored.

As far as output is concerned, decoding theories postulate corresponding encoding process, and abstraction theories necessarily postulate some kind of construction process.

From the experiments that have been reviewed, it is apparent that as far as memory is concerned, there is a fair amount of evidence. Transformational complexity, Yngve complexity, surface structure and predictability (semantically defined) are all important factors in acquisition. Whether they remain important during retention is still an open question.

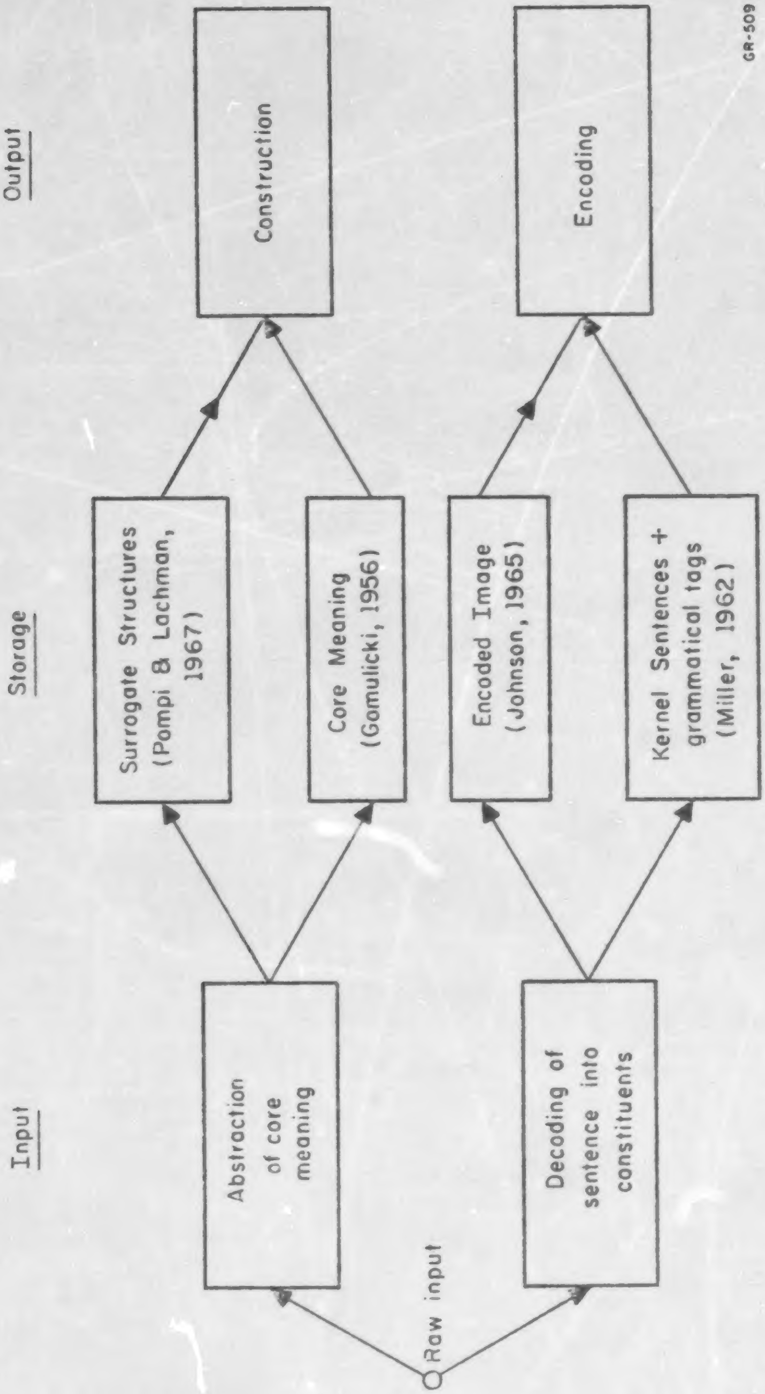

Input
Storage
Output
Raw input
Abstraction of core meaning
Decoding of sentence into constituents
Surrogate Structures (Pompi & Lachman, 1967)
Core Meaning (Gomulicki, 1956)
Encoded Image (Johnson, 1965)
Kernel Sentences + grammatical tags (Miller, 1962)
Construction
Encoding
GR-509

Figure 13. A schematic representation of the kinds of theories discussed herein.

Putting the same point in other words: we have seen that the more predictable the less transformationally complex, and the less complex in terms of its surface structure that a sentence is, the easier it is to learn. We do not know whether or not these same factors affect its retention in the same way. An immediate problem is that we do not know what is stored; the evidence is compatible with several views. These considerations lead to the conclusion that the chief need of the moment is for careful empirical work which will clarify the relationship between these variables, and which will apply them to as yet uninvestigated psychological processes.

THE PRESENT STUDY

At least three kinds of variables are important in the learning and remembering of sentences: predictability, surface structure and transformational complexity. However, theoretical understanding of, and empirical knowledge about these variables are not abundant, at least from a psychological point of view. Neither is there much evidence concerning the relations of these three factors to variables that are known to affect learning and remembering like the length of the retention interval. Nor is there any more evidence about the relations of predictability and surface structure and transformational complexity with one another.

Questions which present themselves include: are predictability and the two complexity variables independent of one another, or do they interact in some fashion, at acquisition, or during retention? Are certain words learned more readily and does this depend on their structural role? Are the associative bonds between words in memory equivalent, or do they differ in strength?

Broad empirical questions of this kind suggest an experiment in which both sentence variables and memory factors (e.g. length of retention interval) are included. Accordingly, a study incorporating some of these variables has been designed and is described below.

A further question arises: how should retention be assessed? There are different ways of measuring retention, and it has been contended that these different methods may not tap the same underlying processes (Adams, 1967). In particular, it has been argued that recall memory and recognition memory are not the same. Whatever their theoretical differences may or may not be, they usually lead to large differences in retention. Since the present study is concerned to establish, if possible, some broad generalizations, rather than clarify minute detail, it seems important to include both recognition and recall.

Recall may be free, or stimulated in some way. Since we are interested in the retention of word strings, primed recall seems to provide a useful method for assessing the strength of various parts of the associative network by priming selected parts of it, and noting which associations are evoked. Presumably the evoked terms will be there with the strongest associative bonds to the word used as a cue or primer. Priming also provides a point of contact with recognition tests as in both the S has a stimulus event at the retention test and the response depends on what has been coded at acquisition, and what remains after the retention interval.

Evidence has been presented which indicates that a difficult problem in research on sentences arises from the fact that different grammatical constructions use different numbers of words. For example, the passive construction normally requires two more words than the equivalent sentence expressed in an active form. Moreover, the number of words in a sentence affects the rate of learning, at least for sentences of seven or less words in length (Martin & Roberts, 1967). This being the case, it was first necessary to determine whether or not the number of words that a sentence contains affected learning when the sentences were of the order of 12 or 13 words in length. A preliminary experiment was carried out to examine the effect of differences in word count and found no differences in retention scores for sentences ranging from 11 to 14 words in length.

Experiment I

This was a recognition experiment. In essence it involved presenting an S with 68 sentences (including four buffer and 64 experimental sentences with two buffers at either end of the list), and then giving him a recognition test in which 66 of the acquisition sentences (made up of 64 experimental and two buffer sentences) were presented, intermixed with 66 new sentences.

Acquisition sentences. These consisted of 64 sentences made up of 8 examples each of 8 different kinds of sentence, plus four additional sentences to provide buffer items at the beginning and end of the list (see Appendix 1). The 8 different sentence types comprised all possible combinations of two levels of the following three variables:

1. Transformational complexity (Tc), or complexity as indexed by the number of transformational operations required in the translating of a base phrase-marker into an actual sentence (see pp.30-31, pp. 37-38 above). The active and the passive were the two operations used. It has been supposed that the passive transformation requires rather more operations than the active (Miller, 1962), and there is some experimental evidence supporting this supposition (pp. 37-38 above). However, although linguists no longer hold to this view, the empirical question remains, and it is not yet settled. So in this study passive sentences indicated high Tc, and active sentences were regarded as possessing low Tc.

2. Sentence predictability (Prd). Predictability may be defined in many ways, and some of them have already been discussed (see pp. 14-22 above). In this experiment prediction was defined in terms of the

real-word probability of the proposition of a sentence. Either the subject and the predicate could provide an unusual combination, or the modifiers and the terms they modified. An example of the first kind of combination is:

(9) The ugly boss was grandly entertained by the poor cleaner with Swiss liqueur.

Phrases which examplify the second kind of incongruence are 'uncouth bishop', 'raucous jury', and 'cocktail banquet'.

None of the sentences were anomalous (Miller & Isard, 1964), in that they did not violate any selectional restrictions, that is, there were no sentences like:

(10) The bishop seduced the stone

or (11) The highway flipped his lid

In the first case, *seduce* is marked $+_{\text{animate}}$ and *stone* is inanimate, and in the second case *highway* is inanimate, and *his* is marked $+_{\text{animate}}$. Although the sentences were unusual, they were not nonsensical. Their unpredictability lay in the fact that their propositions, while possible, were certainly most uncommon. One would expect that predictability, so defined, would correlate quite highly with cloze or successive guessing scores.

3. Surface structure complexity or sentence depth (Yd). Yngve numbers (see pp. 25-28 above) were used to index surface complexity, which was of two levels, simple (low complexity) and complex (high complexity). High Yd or complex sentences had a mean Yngve value of 1.69 (range 1.62 - 1.82), and low Yd or simple sentences had a mean Yngve value of 1.40 (range 1.31 - 1.50).

None of the sentences shared any common content words (nouns, adjectives, verbs and adverbs). Word frequency and sentence length was approximately equal across the 8 groups of 8 sentences. The actual means are shown in Table 6. It can be seen that the passive sentences, on the average, contain about one word more than the active sentences, and that the words comprising the unpredictable sentences are slightly less common than the words from the predictable sentences. With regard to letter count, the active, unpredictable and complex sentences were slightly shorter than the others; mean number of letters for all sentences was 62.5, the range being from 52 to 75.

Distractor sentences. The distractors were obtained by slightly changing the acquisition sentences. Each acquisition sentence gave rise to 8 distractors. In four cases the distractors (or new sentences) were the same as the acquisition (or old) sentences, except that either the subject, object, verb or the noun contained in the adverbial phrase of the sentence was replaced by a word of similar meaning, thus ensuring that no selectional restrictions were violated on account of the substitution.

In the remaining four cases, the new (distractor) sentences differed from the original in that the word order was changed. Either the adverbial phrase was shifted to one of three possible new positions in the sentence, or the adverb associated with the main verb was shifted to the front of the sentence. The adverbial phrase shift was denoted as adjacent, intermediate, or extreme, depending on the magnitude of the change.

TABLE 6

Yngve Value, Content work frequency (from the Thorndike-Lorge (1948) word count), Word count, and Letter count means for the 8 groups of acquisition sentences.

Sentence Type	Mean Yngve value	Mean number of words per sentence	Mean No. of letters per sentence	Mean No. of word of 50/million or greater	frequency of remaining words (per million)
Active Predictable Simple	1.38	12.0	61.9	3.0	18
Active Predictable Complex	1.69	12.1	62.0	3.3	16
Active Unpredictable Simple	1.38	12.1	62.0	2.9	14
Active Unpredictable Complex	1.73	11.8	57.9	3.0	19
Passive Predictable Simple	1.42	12.9	63.9	3.6	16
Passive Predictable Complex	1.69	12.9	61.6	3.5	16
Passive Unpredictable Simple	1.43	13.5	66.9	2.9	14
Passive Unpredictable Complex	1.65	13.0	64.1	2.5	14

Experimental design. All the sentences were learned by every subject; then one group of 72 subjects was given a recognition test immediately after acquisition, while the second group of 48 subjects was tested after a retention interval of 48 hours.

The presentation sequence at acquisition began and finished with two buffer sentences, included to reduce any possible primacy-recency end effects. A sequence for the intervening 64 sentences was obtained by randomly choosing (without replacement) one of the 8 experimental groups (e.g. passive, predictable and simple), randomly choosing a sentence from this group, then repeating the process until every group had been selected once. The process was then repeated to select another eight sentences, and so continued until all 64 sentences had been chosen. Five such sequences were generated, and an equal number of Ss were assigned to each sequence.

For the recognition test, another sequence of acquisition or old sentences was generated, following the rules described in the previous paragraph. Distractor or new sentences were then randomly interleaved between the old sentences with the constraint that no more than three old or new sentences could appear consecutively. A second constraint was that for half the sentences the new version appeared first, and for the remaining sentences the new version appeared second. A third constraint was that at least 35 sentences intervened between the appearances of the new and old versions of a sentence.

Although for every old sentence in a list there was only one distractor, any one of 8 distractors might have been selected. For each S a different kind of distractor appeared with each group of sentences, so that, for example, the 8 active predictable and simple sentences might have as distractors sentences with the verbs changed, while the 8 passive, unpredictable and complex sentences might have as their distractors sentences with their adverbial phrases shifted to an intermediate position. To ensure that each sentence type was paired at least once with each kind of distractor, both of the two groups of Ss were further subdivided into 8 groups, giving 9 Ss per group for the zero retention interval, and 6 Ss per group for the 48 hour retention interval. This between-subjects variable is hereinafter referred to as *Lists*.

1. It had originally been intended to cross distractor type with sentence type, to yield a repeated measures design with the variables Tc, Prd, and Yd, and distractor type completely crossed. Due to an undetected error in the computer program, which generated the stimulus lists, the above design was run instead.

Apparatus.

The experiment was run on the PLATO equipment at the University of Illinois. The PLATO system (as used in this experiment) consisted of 20 terminals linked via interface equipment to a CDC 1604 computer. Each terminal has its own typewriter keyset and CRT display (television screen). The keyset allows the student to send information to the computer, and the television screen presents information prescribed by the computer program. The computer generates visual information on one of two ways: by selecting previously prepared 35 mm slides from a central bank of 122 slides or by plotting characters and figures on the S's individual electronic blackboard. The images from the blackboard and the slide selector are superimposed on the S's television screen.

Each terminal is independent of every other terminal, and each may be individually controlled and monitored by the computer. Thus 20 Ss can be run simultaneously, but independently of one another. All information that is output to and input from an S may be recorded on magnetic tape. A more detailed description of the system may be found in Bitzer, Hicks, Johnson and Lyman (1967), and the references cited by them.

In the present experiment a computer program presented slides containing instructions, presented all the stimuli, and recorded the responses which were entered via the keyset.

Procedure. The experimental procedure consisted of the following steps: after the Ss were seated at their terminals a series of instruction slides were presented (see Appendix 2, slides 1-5). These informed them that they were participating in an experiment on the retention of sentences, and that they were about to write down and learn a set of sentences.

The Ss were then presented with three practice sentences. They were given 31 seconds to write down each sentence on a prepared form which had been placed next to their PLATO keyset. During this time, a slide (see Appendix 2, slide 6) which said "Write down and learn this sentence as quickly as possible" was displayed. Three seconds before the end of the presentation of a sentence, the words "three seconds left" appeared at the bottom of the screen.

The reasons for making an S write out every sentence were twofold. First, and most importantly, to ensure that he paid at least some attention to every word in the sentence, instead of just concentrating on major content words. Second, to prevent (to some degree) the S from using the pattern of words on the screen (each sentence had to be written out over three rows) to help him learn.

Following the practice sentences, more slides were shown (see Appendix 2, slides 8-9). These checked that the Ss understood the procedure, and prepared them for the main acquisition list, comprising the 64 experimental sentences preceded and followed by two buffer sentences. As in the practice trials each sentence was shown for 31 seconds, and the warning "three seconds left" was displayed three seconds before the end of each presentation.

At the retention test (either immediately after or 48 hours later), following more instructions (see Appendix 2, slides 10-16), the Ss were presented with a list of 4 practice sentences composed of two of the practice sentences learned at acquisition and two new (distractor) sentences. Each sentence was shown for 11 seconds, and for each sentence the Ss had to indicate (1) whether or not the sentence was old or new, and (2) how confident they were that their judgment was correct.

The first judgment was made by pressing one of two keys (corresponding in position to Y and U on a normal typewriter) covered by keycaps with the legends "old" and "new" respectively on them. The second judgment was made be pressing one of six keys (corresponding in position to 1,2,3,4,5,and 6 on a normal typewriter) numbered one to six.

During this time a background instruction slide (Appendix 2, slide

17) was shown on the screen. When key "old" or "new" was pressed, the word OLD or NEW appeared on the screen, immediately below the test sentence. When either of keys 1 2,3,4,5, or 6 was pressed the string XXX was written over the appropriate number on slide 17. The process is illustrated in Figure 14.

As at acquisition, three seconds before each trial ended, the warning "three seconds left" was projected at the bottom of the screen.

Following the practice sentences, it was ascertained that the Ss had understood the procedure. Then the recognition test proper began, during which 132 sentences were presented, 66 old sentences and 66 distractors, intermixed as described above in the section on experimental design. One further constraint was that the four buffers were presented first, and then the 128 experimental sentences. The reason for this was to avoid any perturbation due to initial warm up effects, and get the Ss started on the test run.

Results

The Ss judged each sentence to be either "old" or "new", and then expressed the degree of confidence that they had in the correctness of their judgment by giving a rating on a six point scale. If it is the case that Ss are capable of assessing the accuracy of their responses, then utilizing their confidence judgments should provide a more sensitive measure of the strength of the memory trace than is possible with only binary (old/new) responses. For example, a correct judgment about which the S is highly confident would indicate a very well remembered item, but an incorrect judgment about which the S is highly confident would show that the item has been thoroughly forgotten.

Indicate whether this sentence is "old" or "new" by pressing key "old" or key "new."

THE TALKATIVE JANITOR
ACCIDENTALLY SHOOK HIS MOP INTO
THE PROFESSORS FACE

OLD

Indicate how confident you are that your response is correct by pressing one of the keys:

1	2	3	4
(completely certain)	(moderately certain)	(not very certain)	(no idea)

Figure 14. This figure shows the TV screen as it might appear to an S after he had judged the sentence to be old, and judged himself to be moderately certain that his first judgment was correct.

Apart from persuasive independent evidence that Ss are able to evaluate the correctness of their responses with considerable accuracy (Adams, 1967, Wearing (unpublished)), the results of the present experiment also suggest that confidence judgments are related to the strength of the memory trace. For old (previously learned) sentences, it might be expected that correct judgments would be made more confidently than incorrect ones, since the trace of any old sentence would still possess at least some strength, which would mean that an S could not be confident that he had not seen the sentence before, whereas he could be confident that he had. In fact, the mean confidence rating for correct old sentences was 4.74, whereas for incorrect old sentences, it was 4.02. The difference is highly significant.

In the case of new sentences, an S compares the input (a new sentence) against similar (the corresponding old sentence) but slightly different traces. The slight mismatch could be due to either the test sentence being actually new, or to an improperly remembered old sentence. In neither case, however, could a match occur. It follows that the discrepancy in the amount of confidence for correct and incorrect responses should be rather less for new sentences. In fact, for correct new sentences, the mean confidence rating was 4.62, and for incorrect new sentences the mean rating was 4.43. The difference is highly significant.

Response latency may also be regarded as an index of trace strength (Adams, 1967), and if confidence estimates are also a valid estimate of trace strength, then they should be closely related to response latencies. In this case, the latency measure in question is the time delay between the appearance of a sentence on the screen, and the pressing of the "old" or "new" key. In Table 7, the mean latencies for each level of confidence are shown. Clearly correct responses have faster latencies, and the degree of confidence is very closely related to response latency.

For these reasons, it seemed that S's confidence judgments did reflect the strength of the memory trace, and so old/new responses and confidence judgments were combined as shown in Table 8.

In order to have an approximately equal number of cases in each category, categories 2 and 3, and 4 and 5 and 6 were collapsed as shown in Table 8. Each S was then assigned a score for each response, ranging from 1 to 6 (see Table 8), depending on both his old/new response and his confidence estimate. In effect, the binary response was further differentiated in order to provide a more sensitive measure of the strength of the memory trace.

One problem with experiments of this kind is that of response bias, that is, differences in the tendencies of individuals to respond either "old" or "new". One S, for example, may be willing to regard any sentence as old, unless he is absolutely sure that he did not see the sentence before. Another S may adopt the reverse position. However, it is possible to control for the idiosyncratic bias by the use of a repeated measures design which ensures that any particular S's bias is distributed evenly across all levels of the within-subjects variables. In the present experiment, response bias was controlled in this fashion.

The score described in Table 8 were analyzed with a series of

TABLE 7

Latencies for the six different levels of response confidence

Confidence level	0%	20%	40%	60%	80%	100%
Latency	6.84	6.69	6.61	6.34	5.98	5.34

TABLE 8

Distributions of Confidence estimates and the weights assigned to confidence estimates for subsequent analyses.

Response	Incorrect			Correct		
Confidence	High		Low	Low		High
Category	1	2+3	4+5+6	4+5+6	2+3	1
Assigned weights	1	2	3	4	5	6
Frequency	1771	2600	1806	1742	3967	3474

analyses of variance. First the responses to the 64 old sentences were analyzed, then the responses to the new sentences, and finally, analyses were run to determine whether or not the nature of the distractor had any effects on either the recognition of an old sentence or the detection of a new sentence.

The first analysis of variance was carried out on the responses to the old sentences. The completely crossed design had two between-subject factors, retention interval (RI) and Lists (see p. 62 for description), and four within-subject factors, transformational complexity (Tc), Predictability (Prd), Yngve depth (Yd) and replications (rep). The transformed score described in Table 8 provided the dependent variable. The summary table is shown in Table 9.

Of the main effects, RI was highly significant ($p < .001$), as were Prd ($p < .0005$) and Yd ($p < .01$). Neither Tc, Lists, nor Rep was significant. The non-significance of Lists indicates that the recognition of an old sentence was independent of the kind of distractor with which it was paired, that is, sentence type did not interact with distractor type. The non-significance of Rep indicates that there were no overall effects due to list position, it made no difference whether a particular sentence type was in the first, middle, or end position. Interactions with Rep were highly significant in two cases. In the case of Rep x Prd ($p < .001$) the interaction was due to the fact that the difference between predictable and unpredictable sentences in the same list position were different across list positions. That is, the difference in scores between the first predictable and the first unpredictable sentence was not the same as the difference between the third predictable

TABLE 9

Analysis of variance Summary Table for old sentences.

Source	D.F.	Sums of Squares	Mean Squares	F Ratio	Significance Level
RI	1	149.99	149.99	12.07	.001
Lists	7	114.02	16.29	1.31	n.s.
RIxLists	7	63.70	9.10	.73	n.s.
Ss	104	1292.72	12.43		
Tc	1	1.18	1.18	.36	n.s.
RIxTc	1	.12	.12	.04	n.s.
ListsxTc	7	51.35	7.34	2.22	.05
RIxListsxTc	7	19.39	2.77	.84	n.s.
SsxTc	104	343.82	3.31		
Prd	1	41.86	41.86	14.72	.0005
RIxPrd	1	.32	.32	.11	n.s.
ListsxPrd	7	16.45	2.35	.83	n.s.
RIxListsxPrd	7	1.99	.28	.10	n.s.
SsxPrd	104	295.74	2.84		
Yd	1	18.90	18.90	7.17	.01
RIxYd	1	6.40	6.40	2.43	n.s.
ListsxYd	7	23.02	3.29	1.25	n.s.
RIxListsxYd	7	8.02	1.15	.43	n.s.
SsxYd	104	274.14	2.64		
Rep	7	37.11	5.30	1.93	n.s.
RIxRep	7	40.08	5.73	2.08	.05
ListsxRep	49	144.24	2.94	1.07	n.s.
RIxListsxRep	49	104.45	2.13	.77	n.s.
SsxRep	728	2002.42	2.75		
TcxPrd	1	13.75	13.75	6.04	.025
RIxTcxPrd	1	16.24	16.24	7.13	.01
ListsxTcxPrd	7	19.59	2.80	1.23	n.s.
RIxListsxTcxPrd	7	25.46	3.64	1.60	n.s.
SsxTcxPrd	104	236.81	2.28		

TABLE 9 continued

Source	D.F.	Sums of Squares	Mean Squares	F Ratio	Significance Level
TcxYd	1	.20	.20	.07	n.s.
RIxTcxYd	1	.81	.81	.28	n.s.
ListsxTcxYd	7	14.46	2.07	.72	n.s.
RIxListsxTcxYd	7	10.06	1.44	.50	n.s.
*S*sxTcxYd	104	297.82	2.86		
TcxRep	7	20.60	2.94	1.25	n.s.
RIxTcxRep	7	10.34	1.48	.63	n.s.
ListsxTcxRep	49	161.50	3.30	1.40	n.s.
RIxListsxTcxRep	49	109.65	2.24	.95	n.s.
*S*sxTcxRep	728	1719.40	2.36		
PrdxYd	1	2.52	2.52	1.05	n.s.
RIxPrdxYd	1	17.07	17.07	7.15	.01
ListsxPrdxYd	7	22.95	3.28	1.37	n.s.
RIxListsxPrdxYd	7	21.98	3.14	1.31	n.s.
*S*sxPrdxYd	104	248.96	2.39		
PrdxRep	7	86.48	12.35	5.33	.0001
RIxPrdxRep	7	23.00	3.29	1.49	n.s.
ListsxPrdxRep	49	150.30	3.07	1.32	n.s.
RIxListsxPrdxRep	49	161.26	3.29	1.42	n.s.
*S*sxPrdxRep	728	1685.95	2.32		
TcxPrdxYd	1	1.18	1.18	.41	n.s.
RIxTcxPrdxYd	1	6.45	6.45	2.26	n.s.
ListsxTcxPrdxYd	7	24.12	3.45	1.21	n.s.
RIxListsxTcxPrdxYd	7	15.17	2.17	.76	n.s.
*S*sxListsxTcxPrdxYd	104	296.19	2.85		
TcxPrdxRep	7	28.49	4.07	1.73	n.s.
RIxTcxPrdxRep	7	36.20	5.17	2.20	.05
ListsxTcxPrdxRep	49	159.49	3.25	1.39	n.s.
RIxListsxTcxPrdxRep	49	94.57	1.93	.82	n.s.
*S*sxListsxTcxPrdxRep	728	1710.20	2.35		
YdxRep	7	30.42	4.35	1.87	n.s.
RIxYdxRep	7	11.69	1.67	.72	n.s.
ListsxYdxRep	49	131.52	2.68	1.15	n.s.

TABLE 9 continued

Source	D.F.	Sums of Squares	Mean Squares	F Ratio	Significance Level
RIxListsxYdxRep	49	158.71	3.24	1.39	n.s.
SsxYdxRep	728	1696.02	2.33		
TcxYdxRep	7	60.56	8.65	3.90	.0005
RIxTcxYdxRep	7	7.96	1.10	.50	n.s.
ListsxTcxYdxRep	49	180.87	3.69	1.66	n.s.
RIxListsxTcxYdxRep	49	46.50	.95	.43	n.s.
SsxTcxYdxRep	728	1614.33	2.22		
PrdxYdxRep	7	27.75	3.96	1.71	n.s.
RIxPrdxYdxRep	7	13.78	1.97	.85	n.s.
ListsxPrdxYdxRep	49	131.21	2.68	1.16	n.s.
RIxListsxPrdxYdxRep	49	82.49	1.68	.73	n.s.
SsxPrdxYdxRep	728	1685.61	2.32		
TcxPrdxYdxRep	7	32.78	4.68	2.10	.05
RIxTcxYdxPrdxRep	7	5.56	.79	.36	n.s.
ListsxTcxPrdxYdxRep	49	193.85	3.96	1.73	n.s.
RIxListsxTcxPrdxYdxRep	49	112.50	2.30	1.03	n.s.
SsxTcxPrdxYdxRep	728	1621.82	2.23		

and unpredictable sentences. Of course, the order of the sentences was quite arbitrary, so no importance should be attached to the interaction. The same reasoning accounts for the Tc x Yd x Rep interaction, which was also highly significant ($p < .0005$).

The means of the main effects of the various factors are shown in Table 10, and the means for the significant interactions only are shown in Tables 11a, 11b, and 11c.

There was a significant decline in performance over the 48 hour retention period ($p < .001$), unpredictable sentences were recognized better than predictable ones ($p < .0005$), and simple sentences (low Yd) were detected more accurately than complex (high Yd) ones ($p < .01$). The actual magnitude of these differences was very small. The percentage of responses correct for each level of the main effects is shown in Table 12.

Predictability and transformational complexity interacted in that active predictable sentences were not recognized as well as passive predictable sentences, however, the relationship was reversed for unpredictable sentences. Two significant interactions occurred only in the 48 hour retention group, one between Tc and Prd ($p < .01$), and the other between Yd and Prd ($p < .01$).

The second analysis of variance was carried out on the responses to the new sentences. The design was the same as for the first analysis, and the summary table is shown in Table 13 below. Of the main effects, RI was highly significant ($p < .005$), as was Rep ($p < .0001$). Lists was significant to a lesser degree ($p < .025$). None of the other main effects, Tc, Prd, and Yd, attained significance.

TABLE 10

Means for the experimental main effects in experiment II (old sentences).

Variable	Levels	Means	Significance of difference
Retention Interval	0 hrs	4.28	.001
	48 hrs	4.00	
Transformational Complexity	Active	4.16	n.s.
	Passive	4.18	
Predictability	Predictable	4.10	.0005
	Unpredictable	4.24	
Yngve depth	Simple	4.22	.01
	Complex	4.12	
Replications:	1	4.18	n.s.
	2	4.21	
	3	4.14	
	4	4.33	
	5	4.11	
	6	4.17	
	7	4.10	
	8	4.12	

TABLE 11

Means for significant interaction effects among the experimental variables in experiment II (old sentences).

(a)

	Predictable	Unpredictable	Significance
Active:	4.04	4.27	.025
Passive:	4.15	4.21	

(b)

	0 hr retention interval		48 hr retention interval		
	Prd.	Unprd	Prd	Unprd	Significance
Active:	4.20	4.34	3.81	4.17	.01
Passive:	4.23	4.36	4.02	3.98	

(c)

	0 hr retention interval		48 hr retention interval		
	Prd.	Unprd	Prd	Unprd	Significance
Simple:	4.22	4.40	4.08	4.09	.01
Complex:	4.21	4.31	3.76	4.07	

TABLE 12

Percentage of responses correct for the main experimental effects for old sentences.

	0 hour retention	48 hour retention
Active	69.1	62.5
Passive	69.9	62.3
Predictable	68.1	60.2
Unpredictable	70.9	64.7
Simple	69.7	64.5
Complex	69.2	60.4
Grand Mean	69.5	62.4

TABLE 13

Analysis of variance Summary Table for new sentences

Source	D.F.	Sums of Squares	Mean Squares	F Ratio	Significance Level
RI	1	136.50	136.50	8.84	.005
Lists	7	297.81	42.54	2.76	.025
RIxLists	7	73.29	10.47	.68	n.s.
Ss	104	1605.56	15.44		
Tc	1	10.14	10.14	3.12	n.s.
RIxTc	1	.48	.48	.15	n.s.
ListsxTc	7	46.14	6.59	2.03	n.s.
RIxListsxTc	7	62.42	8.92	2.75	.025
SsxTc	104	337.47	3.24		
Prd	1	4.85	4.85	1.48	n.s.
RIxPrd	1	30.53	30.53	9.29	.005
ListsxPrd	7	71.09	10.16	3.09	.005
RIxListsxPrd	7	21.85	3.12	.95	n.s.
SsxPrd	104	3417.01	3.29		
Yd	1	1.84	1.84	.72	n.s.
RIxYd	1	.94	.94	.36	n.s.
ListsxYd	7	39.93	5.70	2.21	.05
RIxListsxYd	7	21.55	3.08	1.19	n.s.
SsxYd	104	268.07	2.58		
Rep	7	179.33	25.62	7.22	.0001
RIxRep	7	98.49	14.07	3.97	.0005
ListsxRep	49	178.03	3.63	1.02	n.s.
RIxListsxRep	49	164.99	3.37	.95	n.s.
SsxRep	728	2582.01	3.55		
TcxPrd	1	2.74	2.74	.91	n.s.
RIxTcxPrd	1	3.65	3.65	1.21	n.s.
ListsxTcxPrd	7	117.33	16.76	5.55	.0001
RIxListsxTcxPrd	7	36.70	5.24	1.74	n.s.
SsxListsxTcxPrd	104	314.22	3.02		
TcxYd	1	16.97	16.97	7.06	.01
RIxTcxYd	1	2.39	2.39	1.00	n.s.
ListsxTcxYd	7	177.08	25.30	10.53	.0001

TABLE 13 continued

Source	D.F.	Sums of Squares	Mean Squares	F Ratio	Significance Level
RIxListsxTcxYd	7	44.59	6.37	2.65	n.s.
SsxTcxYd	104	249.80	2.40		
TcxRep	7	30.58	4.37	1.66	n.s.
RIxTcxRep	7	2.22	.32	.12	n.s.
ListsxTcxRep	49	211.73	4.32	1.64	.005
RIxListsxTcxRep	49	162.99	3.33	1.26	n.s.
SsxTcxRep	728	1919.42	2.64		
PrdxYd	1	14.61	14.61	5.00	.05
RIxPrdxYd	1	2.09	2.09	.71	n.s.
ListsxPrdxYd	7	217.08	31.01	10.62	.0001
RIxListsxPrdxYd	7	51.39	7.34	2.51	n.s.
SsxPrdxYd	104	303.78	2.92		
PrdxRep	7	22.99	3.28	1.25	n.s.
RIxPrdxRep	7	19.81	2.83	1.08	n.s.
ListsxPrdxRep	49	137.87	2.81	1.07	n.s.
RIxListsxPrdxRep	49	112.64	2.30	.88	n.s.
SsxPrdxRep	728	1910.78	2.62		
TcxPrdxYd	1	3.63	3.63	1.58	n.s.
RIxTcxPrdxYd	1	3.94	3.94	1.71	n.s.
ListsxTcxPrdxYd	7	134.76	19.25	8.36	.0001
RIxListsxTcxPrdxYd	7	13.45	1.25	.83	n.s.
SsxTcxPrdxYd	104	239.42	2.30		
TcxPrdxRep	7	33.29	4.76	1.88	n.s.
RIxTcxPrdxRep	7	17.01	2.43	.96	n.s.
ListsxTcxPrdxRep	49	277.48	5.66	2.24	.0001
RIxListsxTcxPrdxRep	49	133.93	2.73	1.08	n.s.
SsxListsxTcxPrdxRep	728	1840.71	2.53		
YdxRep	7	57.54	8.22	3.28	.005
RIxYdxRep	7	34.21	4.89	1.95	n.s.
ListsxYdxRep	49	152.73	3.12	1.25	n.s.
RIxListsxYdxRep	49	171.46	3.50	1.40	.05
SsxYdxRep	728	1821.83	2.50		

TABLE 15 continued

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	Significance Level
TcxYdxRep	7	49.61	7.09	2.74	.01
RIxTcxYdxRep	7	17.87	2.55	.99	n.s.
ListsxTcxYdxRep	49	276.48	5.64	2.18	.0001
RIxListsxTcxYdxRep	49	167.55	3.42	1.32	n.s.
SsxTcxYdxRep	728	1885.72	2.59		
PrdxYdxRep	7	26.91	3.84	1.52	n.s.
RIxPrdxYdxRep	7	27.47	3.92	1.56	n.s.
ListsxPrdxYdxRep	49	242.60	4.95	1.96	.0005
RIxListsxPrdxYdxRep	49	99.80	2.04	.81	n.s.
SsxPrdxYdxRep	728	1836.82	2.52		
TcxPrdxYdxRep	7	32.25	4.61	1.75	n.s.
RIxTcxPrdxYdxRep	7	8.55	1.22	.46	n.s.
ListsxTcxPrdxYdxRep	49	239.27	4.88	1.85	.0005
RIxListsxTcxPrdxYdxRep	49	118.23	2.41	.90	n.s.
SsxTcxPrdxYdxRep	728	1941.91	2.61		

Several interactions involving Lists were highly significant. In general, such interactions indicate that distractor type interacted with sentence type, in other words, whether or not a new sentence was detected jointly depended both on the kind of sentence that is was, and the distractor form in which it was expressed. The means for the main effects of the various variables are shown in Table 14, and the means for the significant interaction are shown in Tables 15a, 15b, 15c and 15d.

As with the old sentences, detection of new sentences becomes less accurate over the 48 hour retention interval. However, sentence type has no effect on detectability. From Tables 14 and 15d, it can be seen that the detection of new sentences improves throughout the course of the list. For the immediate retention groups, the improvement starts straight away, and for the 48 hour retention groups it begins about half-way through the list. No such effect occurred with the old sentences.

Predictability interacted strongly with retention interval, and weakly with Yngve depth, but Yngve depth did interact quite strongly with transformational complexity. Unpredictable sentences were difficult to detect both after 48 hours, and when they were complex (high Yd). Active-complex and Passive-simple sentences were detected more easily than either Active-simple or Passive-complex sentences, and this result parallels one obtained by Martin & Roberts & Collins (1968).

The absolute differences were very small, the percentage of responses correct for each level of the main effect is shown in Table 16.

In order to look directly at the effects of distractor type on recognition, two more analyses were carried out. First the effect of

TABLE 14

Means for the experimental main effects in experiment II (new sentences).

Variable	Levels	Means	Significance of difference
Retention interval	0	3.72	.005
	48	3.45	
Transformational complexity	Active	3.42	n.s.
	Passive	3.48	
Predictability	Predictable	3.59	n.s.
	Unpredictable	3.64	
Yngve depth	Simple	3.60	n.s.
	Complex	3.63	
Replications	1	3.65	0001
	2	3.41	
	3	3.40	
	4	3.50	
	5	3.63	
	6	3.74	
	7	3.73	
	8	3.88	

TABLE 15

Means for the significant interaction effects among the experimental variables in experiment II (new sentences).

(a)

	Predictable	Unpredictable	Significance
0 hr retention	3.65	3.80	.005
48 hr retention	3.50	3.40	

(b)

	Active	Passive	Significance
Simple	3.52	3.68	.01
Complex	3.64	3.62	

(c)

	Predictable	Unpredictable	Significance
Simple	3.53	3.67	.05
Complex	3.65	3.61	

(d)

	Replications								Significance
	1	2	3	4	5	6	7	8	
0 hr retention	3.93	3.53	3.56	3.70	3.64	3.74	3.81	3.88	.0005
48 hr retention	3.24	3.23	3.15	3.20	3.63	3.74	3.63	3.78	

TABLE 16

Percentage of responses correct for the main experimental effects for new sentences.

	0 hour retention	48 hour retention
Active	54.1	48.7
Passive	56.3	50.4
Predictable	53.6	51.2
Unpredictable	56.9	47.9
Simple	54.2	49.6
Complex	56.2	49.5
Grand Mean	55.2	49.5

distractor type on the detectability of old sentences with which the particular kinds of distractors were associated was examined, and then the detectability of the distractors (or new sentences) themselves, as a function of which of the 8 kinds of distractors they represented, was investigated. In the analyses of variance, the design was the same as for the previous analyses, except that three within subject factors, Tc, Prd and Yd were replaced by Distractor type (Dis). For the first analysis, that of old sentences, each sentence is categorized according to the distractor type of its corresponding new sentence. For the second analysis, that of new sentences, each sentence is categorized according to the type of distractor that it actually was.

The analysis of variance summary table for the old sentences is not shown. The main effect that is of interest, distractor type, is not significant. Apparently, the type of distractor associated with an old sentence made little difference to whether or not it would be correctly recognized.

Of rather more consequence are the results for the new sentences, the distractors themselves. The analysis of variance summary table is shown in *Table 17*. Many of the source lines are the same as in *Table 13*. The new main effect, distractor type is highly significant ($p < .0001$). Rep is signifcant ($p < .0001$), and there are several significant interactions. The means for the main effects, and interactions, are shown in *Table 18a and b*, as well as *Figure 15*. There is a significant Dis x Rep interaction which indicates that sentence type and distractor type interacted together. The interaction between Lists and Distractors

TABLE 17

Analysis of variance Summary table for the effect of distractor types in new sentences.

Source	D.F.	Sums of Squares	Mean Squares	F Ratio	Significance Level
RI	1	136.50	136.50	8.83	.005
Lists	7	297.81	42.54	2.76	.025
RIxLists	7	73.29	10.47	.68	n.s.
Ss	104	1605.56	15.44		
Dis	7	300.08	42.87	15.19	.0001
RIxDis	7	90.62	12.95	4.59	.0001
ListsxDis	49	558.11	11.39	4.04	.0005
RIxListsxDis	49	205.31	4.19	1.48	.025
SsxDis	728	2054.47	2.82		
Rep	7	179.33	25.62	7.22	.0001
RIxRep	7	98.49	14.07	3.97	n.s.
Listsx Rep	49	178.03	3.63	1.02	n.s.
RIxListsxRep	49	164.99	3.37	.95	n.s
SsxRep	728	2582.01	3.55		
DisxRep	49	239.63	4.89	1.89	.0005
RIxDisxRep	49	111.54	2.28	.88	n.s.
ListsxDisxRep	343	1551.74	4.52	1.75	.0001
RIxListsxDisxRep	343	982.25	2.86	1.11	n.s.
SsxDisxRep	5096	13157.18	2.58		

TABLE 18

Main interaction effects for distractor type

(a)

Distractor type

Word change				Position shift				Significance
Subject	Verb	Object	Adv. Phrase	Adjacent	Intermediate	Extreme	Adverb	
3.91	3.27	3.82	3.51	3.54	3.57	3.98	3.61	.0001

(b)

Distractor type

	Word change	Position shift	Significance
0 hr retention	3.58	3.87	.0001
48 hr retention	3.21	3.39	

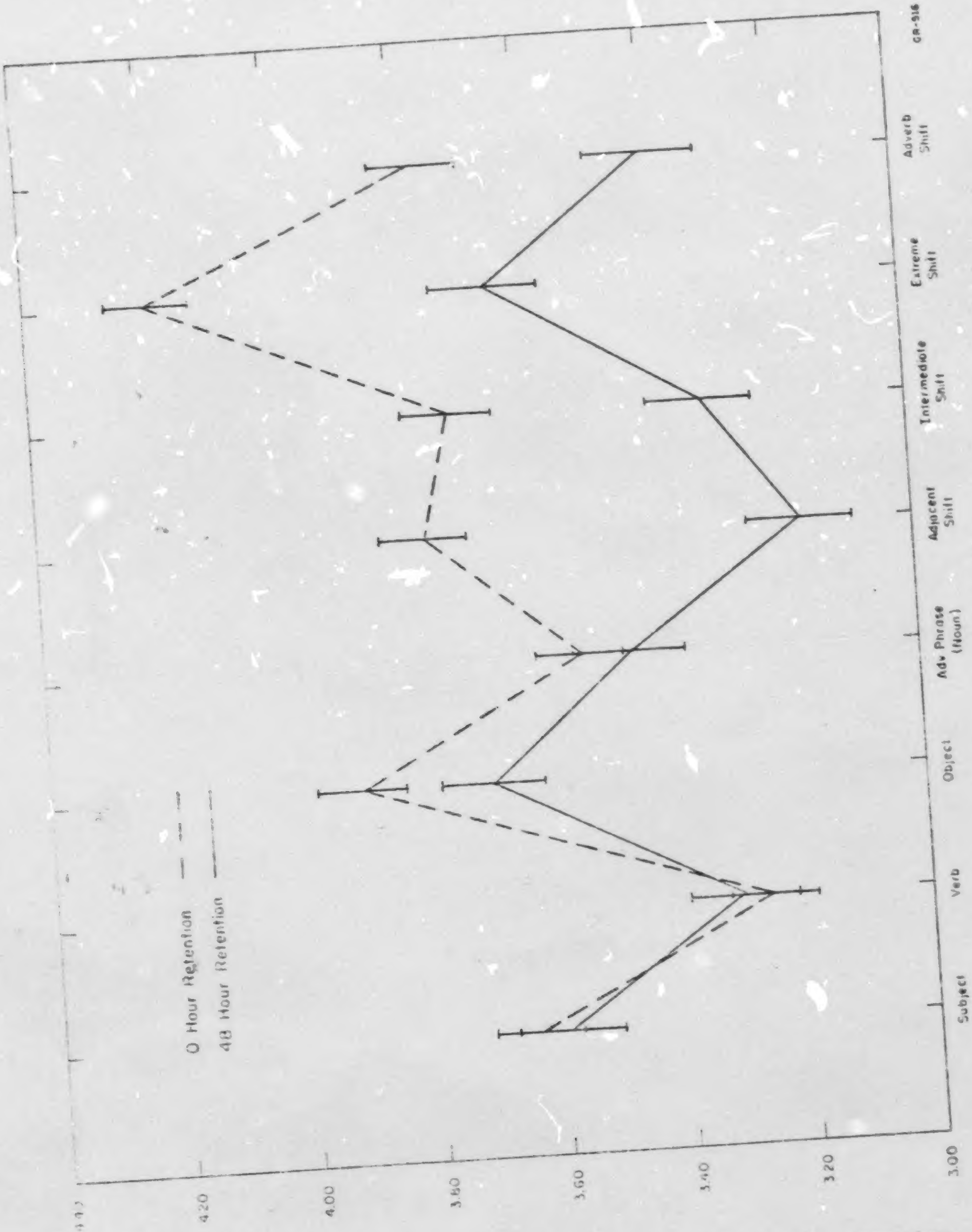


Figure 15. Detection scores for new sentences for each retention interval and distractor type group.

indicates the same phenomenon.

Overall, the significant effect of distractor type is due to the poor detectability of changes in the verb, and the high detectability of changes in the object and in extreme shifts in the position of the adverbial phrase. However there is a strong interaction with retention interval. After 48 hours there is little decline in the detectability of distractors with word change, but a considerable and significant decline in the detectability of word and phrase position changes ($p < .0001$).

The significant variation due to replications suggest that contrary to the case of old sentences, the detectability of new sentences improved throughout the list. However this phenomenon was not consistent for all distractors. Again the interaction is probably due to unique joint effects of sentence and distractor types.

In Table 19, the means for the Distractor type x Lists interactions are shown. In addition, the sentence type is also indicated. An examination of this table fails to yield any systematic relationship; certainly no strong relationship between particular sentence tyep and distractor type combinations appear.

Experiment II

In this experiment cues were used to prime recall. It was similar to the second experiment in that the Ss were presented with 68 sentences including 4 buffer sentences during acquisition, but for the retention test the Ss were given cues, or primers, and asked to recall

TABLE 19

Means for the interaction of distractor types with Lists in new sentences. (Sentence type is also shown).

Distractor type	Lists								Row means
	1	2	3	4	5	6	7	8	
Subject	3.08	3.77	3.77	4.36	4.03	3.35	3.15	3.43	3.62
Verb	3.49	2.98	3.11	3.39	3.77	3.14	3.00	3.28	3.27
Object	3.45	3.81	4.17	4.09	4.32	3.38	3.10	4.17	3.82
Adv. P	3.67	3.60	3.23	3.50	3.95	3.06	3.31	3.71	3.50
[illegible]	3.72	3.29	3.64	3.51	3.96	4.23	2.92	3.04	3.54
[illegible]	3.07	3.56	3.75	3.98	3.98	3.30	3.57	3.34	3.57
[illegible]	[illegible]	4.15	3.74	4.17	3.83	4.27	3.33	4.23	3.98
[illegible]	[illegible]	3.65	4.07	3.53	3.55	3.36	3.52	3.56	3.61

Key to sentence type:

1 = active,	predictable,	simple
2 = active,	predictable,	complex
3 = active,	unpredictable,	simple
4 = active,	unpredictable,	complex
5 = passive,	predictable,	simple
6 = passive,	predictable,	complex
7 = passive,	unpredictable,	complex
8 = passive,	unpredictable,	simple

the remainder of the sentences.

Acquisition sentences. These were exactly the same as the acquisition sentences used in experiment I.

Priming cues. There were four possible primers for each sentence; either the logical subject (plus adjective), verb (plus adverb), logical object (plus adjective) or the noun (plus adjective) from the adverbial phrase.

Experimental design. All the sentences were learned by every S; then one group of 36 Ss was given a recall test immediately, while a second group of 36 Ss was tested after a retention interval of 48 hours. The acquisition phase was exactly the same as that of experiment I.

For the retention test, a further constrained random order of the sentences was generated and then the cue words were taken from these sentences for presentation to the Ss. Since each sentence possessed four possible cues, four lists of cues were constructed, so that each sentence and each cue type appeared together exactly once. For the sake of convenience, this variable is referred to as Lists. Both groups of 36 Ss were subdivided into four sub groups of 9 Ss each, and at recall each sub-group was given a different list.

Apparatus. As in the other two experiments, the PLATO equipment was used.

Procedure. The procedure at acquisition was the same as for experiment I. At retention, the Ss were instructed that they would be presented with two cue words, and that they then had to write out the rest of the sentence (Appendix 2, slides 18-22). They were then shown a practice series of 3 cue word pairs for 35 seconds each, dur ing which time they attempted to recall and write down the appropriate sentence. During this time, a background slide was displayed (Appendix #, slide 23). As at acquisition, three seconds before each trail ended, the warning "three seconds left" was projected at the bottom of the screen. Following the practice items, it was ascertained that the Ss understood the procedure. Then, using the same procedure as for the practice items, the main recall test began, during which 64 sets of cues were presented to each S.

Results

For each sentence the number of words correctly recalled (without regard to order) was scored. There were two sentences in each of the 8 sentence-type x 4 cue-type categories, and the number of words correct (excluding the two cue words) in both sentences, were added together, to give 32 scores per S.

Both the scoring method and the combining of scores was made necessary because of the low level of recall. In order that the cell entries were not preponderantly zero, it was first necessary to give an S credit for any word that he remembered correctly and then it was still necessary to combine sentences in the manner described above. Even so, the mean number of words correct for both sentences added together was only 4.64, or just over two words per sentence.

It is possible that in scoring only the number of words correct one penalizes the S who recalls the core meaning of the sentence or phrase, but does not use the original words to express this meaning. In order to test that possibility the recall protocols were rated on a four point scale. Each sentence contained four clearly definable content phrases, the subject, verb, object and adverbial phrase. For each phrase whose appropriate meaning was recalled, the S received one point. If the approximate meaning of the whole sentence was recalled, the S received 4 points.

It was found that a S usually clearly recalled a phrase, or he did not, there were only a negligible number of doubtful cases. Two judges scored the protocols, in all but a few cases (less than 1%), there was no disagreement. The relationship between the rated score and the number of words correct was extremely strong, making these results consistent with that of Cofer (1941). The mean number of words correct for each rating category and their distribution is shown in Table 20.

Because of the high correlation of the two sets of scores, the number of words correct was preferred as the more sensitive measure.

These scores were analyzed with a 7-way analysis of variance. The design had two between subject factors, retention interval (RI) and Lists, and four within subject factors, transformational complexity (Tc), Predictability (Prd), Yngve depth (Yd) and Cue type (Cues). The design was completely crossed. The summary table is shown in Table 21.

Of the main effects, Prd ($p < .0001$), Yd ($p < .0001$) and Cues ($p < .0001$) were highly significant, Tc was weakly significant ($p < .05$), while RI and Lists were not significant.

Although Lists itself was not significant, it did enter into some highly significant interactions. These interactions indicate that to some degree cues and sentences combine together non-additively. In

TABLE 20

Mean number of words per sentences correct for each rating category and the proportional distribution of the number of words correct per sentence within each category.

Number of words Correct per sentence	Proportion of words correct per sentence within each category for each level of number of words correct			
	Rating Categories			
	1	2	3	4
0	.50	.01		
1	.31	.09		
2	.16	.34	.09	
3	.03	.39	.18	01
4		.12	.32	.06
5		.04	.23	.10
6		.01	.12	.17
7			.06	.26
8				.22
9				.09
10 (or more)				.09
	1.00	1,00	1,00	1,00
Mean number of words correct per sentence.	.72	2,68	4 29	7.11

TABLE 21

Analysis of Variance Summary Table for Experiment III

Source	D.F.	Sums of Squares	Mean Squares	F Ratio	Significance Level
RI	1	632.31	632.31	3.53	n.s.
Lists	3	334.57	111.52	.62	n.s.
RIxLists	3	618.70	206.23	.12	n.s.
Ss	64	11467.79	179.18		
Tc	1	38.80	38.80	4.02	.05
RIxTc	1	2.30	2.30	.02	n.s.
ListsxTc	3	188.66	62.89	6.52	.001
RIxListsxTc	3	9.90	3.30	.34	n.s.
SsxTc	64	617.44	9.65		
Prd	1	492.28	492.28	41.52	.0001
RIxPrd	1	105.49	105.49	8.90	.005
ListsxPrd	3	79.33	2.64	2.23	n.s.
RIxListsxPrd	3	15.40	5.01	.42	n.s.
SsxPrd	64	758.76	11.86		
Vd	1	266.10	266.10	39.69	.0001
RIxVd	1	7.45	7.45	.11	n.s.
ListsxVd	3	25.88	8.63	1.29	n.s.
RIxListsxVd	3	87.04	2.90	4.33	.01
SsxVd	64	429.07	6.70		
Rep	3	750.92	250.31	20.18	.0001
RIxRep	3	124.33	41.44	3.34	.05
ListsxRep	9	243.25	27.03	2.18	.05
RIxListsxRep	9	39.45	4.38	.35	n.s.
SsxRep	192	2381.03	12.40		
TcxPrd	1	.10	.10	.01	n.s.
RIxTcxPrd	1	37.26	37.26	3.71	n.s.
ListsxTcxPrd	3	92.04	30.68	3.05	.05
ListsxRIxTcxPrd	3	23.05	7.68	.76	n.s.
SsxTcxPrd	64	643.08	10.05		
TcxVd	1	24.38	24.38	2.83	n.s.
RIxTcxVd	1	16.17	16.17	1.87	n.s.

TABLE 21 continued

Source	D.F.	Sums of Squares	Mean Squares	F Ratio	Significance Level
ListsxTcxYd	3	45.05	15.02	1.74	n.s.
RIxListsxTcxYd	3	76.31	25.44	2.95	.05
SsxTcxYd	64	552.00	8.62		
TcxRep	3	167.07	55.69	5.33	.005
RIxTcxRep	3	31.89	10.63	1.02	n.s.
ListsxTcxRep	9	405.64	45.07	4.31	.0001
RIxListsxTcxRep	9	139.32	15.48	1.48	n.s.
SsxTcxRep	192	2006.15	10.45		
PrdxYd	1	343.02	343.02	22.34	.0001
RIxPrdxYd	1	37.26	37.26	2.43	n.s.
ListsxPrdxYd	3	7.31	2.44	.16	n.s.
RIxListsxPrdxYd	3	14.15	4.72	.31	n.s.
SsxPrdxYd	64	982.79	15.36		
PrdxRep	3	72.29	24.10	2.17	n.s.
RIxPrdxRep	3	14.80	4.93	.44	n.s.
ListsxPrdxRep	9	865.72	96.19	8.66	.0001
RIxListsxPrdxRep	9	134.90	14.99	1.35	n.s.
SsxPrdxRep	192	2131.50	11.10		
TcxPrdxYd	1	62.34	62.34	6.23	.05
RIxTcxPrdxYd	1	8.88	8.88	.89	n.s.
ListsxTcxPrdxYd	3	148.13	49.38	4.97	.005
RIxListsxTcxPrdxYd	3	133.97	44.66	4.50	.01
SsxTcxPrdxYd	64	635.33	9.93		
TcxPrdxRep	3	146.42	48.81	4.89	.01
RIxTcxPrdxRep	3	129.75	43.25	4.34	.05
ListsxTcxPrdxRep	9	266.89	29.65	2.97	.005
RIxListsxTcxPrdxRep	9	136.49	15.17	1.52	n.s.
SsxTcxPrixRep	192	1915.50	9.98		
YdxRep	3	111.33	37.11	2.99	.05
RIxYdxRep	3	23.68	7.89	.64	n.s.
ListsxYdxRep	9	357.75	39.75	3.20	.005
RIxListsxYdxRep	9	146.94	16.33	1.32	n.s.
SsxYdxRep	192	2382.86	12.41		

TABLE 21 continued

Source	D.F.	Sums of Squares	Mean Squares	F Ratio	Significance Level
TcxYdxRep	3	7.25	2.42	.27	n.s.
RIxTcxYdxRep	3	52.61	17.54	1.92	n.s.
ListsxTcxYdxRep	9	113.49	12.61	1.38	n.s.
RIxListsxTcxYdxRep	9	101.75	11.31	1.24	n.s.
SsxTcxYdxRep	192	1749.59	9.11		
PrdxYdxRep	3	49.10	16.37	2.12	n.s.
RIxPrdxYdxRep	3	22.56	7.52	.97	n.s.
ListsxPrdxYdxRep	9	319.47	35.50	4.60	.0001
RIxListsxPrdxYdxRep	9	80.14	8.91	1.15	n.s.
SsxPrdxYdxRep	192	1482.80	7.72		
TcxPrdxYdxRep	3	60.00	20.00	2.30	n.s.
RIxTcxPrdxYdxRep	3	32.48	10.83	1.25	n.s.
ListsxTcxPrdxYdxRep	9	196.86	21.87	2.52	.01
RIxListsxTcxPrdxYdxRep	9	45.40	5.04	.58	n.s.
SsxTcxPrdxYdxRep	192	1669.65	8.70		

other words, a particular cue depends for its effect, not only on the sentence type (active, predictable or whatever), but also on the unique characteristics of particular sentences. The implications of the interactions involving Lists is not entirely clear, but where they exist they do serve notice that a particular relationship depends in significant part on properties of the sentence other than those measured by the experimental variables.

Another characteristic of the analysis summarized in Table 21 is the very large variance due to *Ss*. Variation between *Ss* accounted for 28% of the total variance, as against 4% for the old and 7% for the new sentences in experiment I. One consequence of this situation is that retention interval, a between subject variable, is not significant, even though the difference between the means is comparable with other mean differences. Another possible consequence is an interaction between Lists and *Ss*. Because *S* was nested in Lists, the existence of such an interaction cannot be tested, but it seems likely that at least some of the differences ostensibly due to Lists may have actually been due to *Ss*. The means of the main experimental effects, and their interactions are shown in Tables 22, 23, and 24.

Unpredictable sentences were learned and retained to a higher level than predictable ones, and simple sentences were learned and retained more effectively than complex ones. Of the cues, the object was most effective, followed by the subject, adverbial phrase and verb in that order. The verb was significantly inferior to the other three cues ($p < .001$), and the object was more effective than the next best

TABLE 22

Means for experimental main effects, and interactions with retention interval.

Retention Interval	Te		Predictability		Yd		Cues				Row Means
	Act	Pass	Prd	Unprd	Simp	Comp	Subj	Verb	Obj	Adv P	
0 hrs	5.02	5.30	4.49	5.84	5.36	4.77	5.35	3.87	6.18	5.15	5.16
48 hrs	4.00	4.24	3.87	4.36	4.40	3.83	4.42	3.57	4.49	3.99	4.12
Col means	4.51	4.77	4.18	5.10	4.88	4.30	4.89	3.72	5.34	4.57	
Significance level (main effects)	.05		.0001		.0001		.0001				
Significance level (interaction)	n.s.		.005		n.s.		.05				

TABLE 23

Means for the interaction of Cue type with transformational complexity, predictability, and Yngve depth.

	Tc		Predictability		Yd	
	Active	Passive	Prd	Unprd	Simple	Complex
Subject	4.82	4.94	4.45	5.34	5.27	4.50
Verb	3.65	3.92	3.47	4.08	3.74	3.80
Object	4.79	5.88	5.00	5.67	5.81	4.86
Obj. P	4.77	4.36	3.81	5.32	5.10	4.03
Significance Level (interaction)	.004		n.s.		.05	

TABLE 24

Means for the Yngve depth x Predictability interaction

	Predictable	Unpredictable
Simple	4.13	5.82
Complex	4.22	4.38

Significance Level = .0001

cue, the subject ($p < .05$). The difference between the subject and the adverbial phrase was not significant.

These interactions among the experimental variables are large enough to deserve mention. Predictability interacted with retention interval ($p < .005$), Tc with cue type ($p < .005$) and Yd with predictability ($p < .05$).

The interaction of Prd with RI is due to the fact that Prd had a much greater effect at immediate recall than 48 hours later, although the difference is still significant for the later interval ($p < .001$). The interaction of Tc with cue type is due to the significantly greater efficacy of the object in passive as opposed to active sentences ($p < .001$). The interaction of Prd with Yd is due to the greater recall of simple unpredictable sentences.

DISCUSSION

The aim of these two experiments was to examine the relationship between three sentence variables (transformational complexity (Tc), Predictability (Prd), and surface structure complexity or Yngve depth (Yd)), and retention. In the first experiment, retention was assessed by a recognition method, and in the second experiment, it was measured by cued or primed recall.

Effect of sentence type on retention

Recognition. In the first experiment, high Prd and low Yd (simple sentences) resulted in significantly better recognition of old sentences at both the zero and 48 hour retention intervals, even though the overall level of retention declined significantly over the 48 hour period. Transformational complexity was not significant, active and passive sentences being recognized equally well.

The detection of new sentences did not depend on either Tc, Prd or Yd, but rather on the nature of the distractor type of the new sentence and the length of the retention interval. However, the detectability of low Prd new sentences, declined significantly more over the 48 hour retention period than high Prd sentences, a result paralleled by a non-significant trend in the old sentences.

The distractors were obtained by either a word change (replacing the logical subject, verb, logical object, or the noun from the adverbial phrase with an appropriate synonym) or a position shift (moving the verb's adverb to the front of the sentence, or changing the adverbial phrase to an adjacent, intermediate or extreme position in the sentence). The distractor type had a considerable effect on the detection of new sentences. The object was associated with significantly higher, and the verb with significantly lower detection scores, with the subject and adverbial phrase noun being intermediate. The effect of position shift was a simple function of the amount of shift, with adjacent shifts being the least, and the extreme shifts being the most detectable. There was a significant interaction between retention interval and distractor type. The detectability of word changes did not change over the 48 hour retention interval, whereas the detectability of position shifts declined markedly, particularly in the case of the adjacent shift.

Cued recall. In the cued recall experiment, high Prd and low Yd resulted in significantly better recall at both the zero and the 48 hour retention intervals, even though the over all level of retention declined over the 48 hour period. Passive sentences were recalled better than active ones, but the significant interaction of Tc X Lists makes the reality of this result rather questionable.

Unpredictable sentences were recalled much better at the zero retention interval than after 48 hours, even though the effect of Prd was still evidence then. This result parallels the outcome of the recognition experiment Another result which parallels one from first experiment is the relatively high level of recall of simple unpredictable sentences compared with complex unpredictable ones.

The effect of cue type (logical subject, verb, logical object or the noun from the adverbial phrase) was highly significant. The object was the most effective cue, and the verb the least effective with the subject and the adverbial phrase noun being intermediate. Over the 48 hour retention interval, the differences between these cues fall away to some degree, the decline in effectiveness as a cue being less for the subject than for the object. Cue type interacts with Tc in that the object in a passive sentence is a highly effective cue whereas the noun from the adverbial phrase is much less effective in a passive than in an active sentence.

Theories of memory of sentences

In the introduction, a number of theories of memory of sentences were outlined, and we now turn to the question of the extent to which the evidence of these present experiments supports them.

Surrogate structure theory. There were four families of theories discussed. The first is exemplified by Pompi & Lackman (1967) who proposed that what is stored are surrogate structures, based (in part) on the word order of the original text, and the associative relations of those words, but not necessarily containing any copies of representations of either the actual words or the grammatical structure linking them (pp 34-35 above).

If this theory holds, then one might expect that word change distractors would be detected less readily than position shift distractors, since the former would not be semantically inconsistent with the remaining words (thus preserving the associative relations), while the latter change word order, which information is retained to some degree in the surrogate structure. Position shifts are certainly detected more readily immediately after learning, but 48 hours later there is no difference between word change and position shift distractors, depending on the amount of shift, no differences would be predicted between the word change distractors. The predicted differences did arise in the first case, but there were significant differences amongst the word change distractors as well. If the surrogate structures are actually formed, then they are independent of neither the actual words in the text, nor their grammatical role.

Structural variables such as Tc and Yd should also not be effective-but clearly surface structure complexity exerts a considerable influence on learning. If associative relations are important, then unpredictable sentences should be more difficult to learn, but this was not the case, in fact, the reverse was true.

More evidence which tells against the Pompi-Lachman position comes from the finding that the number of words correct in the cued recall experiment was strongly related to the amount of sentence content recalled. If the sentences were stored as surrogate structures then one would not expect that recall would be necessary in terms of the original words of the sentences, but that paraphrase would be as frequent as verbatim reproduction.

Surrogate structure theory derives from Gestalt theory, and perhaps more immediately from Bartlett (1932). It seems clear that it is inadequate, principally for the reason that its forebears failed--it lacks a mechanism for handling fixed traces and the relation between them.

Mnemic Abstraction theory. The second family of theories that we considered tries to meet the fixed trace problem of the Gestalt theory, yet preserve its dynamic flexibility by positing some mechanism of abstraction which takes and stores the core meaning of a sentence. Gomulicki (1956) has offered the most precise statement of this type of theory, claiming that an S omits unimportant words and phrases so that what is remembered is an action-agent-effect unit.

The present results run somewhat counter to those of Gomulicki. He found that verbs were best remembered, followed by agents (the logical subjects), followed by the recipient of an action (the logical object). In the first experiment, verbs were detected least readily, followed by subjects, with objects being detected best of all. The same relation held in the second experiment, where objects provided the most effective cue, and verbs the least effective.

An abstraction theory would predict that adverbial phrases would be the most poorly remembered, since they are external to the action-agent-effect unit, but both the detectability and the cue value of the object were better than those of the verb, and almost as good as those of the subject. Another prediction from abstraction theory is that word changes should be more detectable than position shifts, since the action-agent-effect unit should depend more on the actual words than on their order, but his prediction was not confirmed by the first experiment.

Deep structure theory. A third theory comes from transformational grammar. In one version (Miller, 1962), it is argued that an important determinant of learnability is the number of transformations intervening between the base structure of a sentence (which is stored) and its surface structure. A kernel sentence (the core meaning) is stored plus grammatical tags which indicate whether or not the sentence is passive, thelocation of the adverb etc. These two components are stored separately.

The present experiments do not support these notions; active and passive sentences (the latter, according to Miller, requiring one more intervening transformation) being equally easy to recognize, and passive sentences being rather easier to recall. If the semantic and syntactic information is stored separately, then variables affecting them should not interact, but predictability (a semantic variable) does interact with both Tc and Yd, two syntactic variables. Moreover, Miller's theory provides no means of explaining the differences in performance at the zero and 48 hour retention intervals.

The Miller view has been criticized on other grounds (pp. 38-39 above) and a more general and somewhat different version has been put forward by Fodor & Garrett (1966). They argue that what is stored is some representation of the deep structure, and the important variable is not the number of transformations intervening between the deep structure of a sentence, but the clues to deep structure that are present in the surface structure. In addition, they agree with Chomsky (1965) that all meaning bearing elements are present in the base, among which is the passive.

Such a view would predict that word position shifts would be more poorly detected than word changes since the former involve optional transformations that are not relevant to the structure of the base, while the latter involve elements of the base. The results of the first experiment contradict this prediction.

It would be also predicted that sentences with a complex surface structure should be more difficult to learn because the number of transformations intervening between the surface and deep structures is greater. In both experiments, high Yd sentences were the more difficult to learn.

Surface structure theory. A fourth type of theory emphasizes the importance of surface structure. Proponents of this position (e.g. Johnson, 1965) usually assert that what is stored are chunks whose boundaries are derived from the surface structure of the sentence. Chunks may, of course, be nested within one another as in tree diagrams.

A variation on this view comes from Martin, Roberts & Collins(1968) who claim that when a sentence is learned, its grammatical structure is first determined (a task that varies in difficulty according to the surface structure complexity of the sentence), and then word classes are differentially processed into memory, depending on their importance for the sentence.

It seems reasonable to suppose that such a theory would predict that distractors which involve rearranging chunk order (word position shifts) would be easier to detect than distractors which merely changed words, and did not interfere with the chunk structure. That in fact, was the case for immediate retention. However, after 48 hours the differences between the two types of distractors had vanished, which suggests that chunks based on surface structure are not critical for LTM.

Complexity of surface structure was predictably important (simple sentences were learnt more readily than complex ones), but the theory does not have any means for dealing with either semantic variables, or

STM - LTM differences, both fairly serious drawbacks. However, it does provide a partial account of the differences that were found between the cues and the word change distractors. If what is stored are nested chunks, following the tree diagram of the original sentence, then it might be reasonably expected that the associative bonds between the chunks were a function of the closeness of their tree links. It is clear that in the case of a simple type of tree, the object is linked most closely to the verb on one side, and an adverbial phrase on the other, whereas all the other main constituents have only one such close linkage. Nevertheless, this account is insufficient, as it does not account for the inferiority of the verb both as a distractor and a cue, nor does it handle the situation where the tree is differently structured.

Some problems and theoretical suggestions

It can be seen that although none of these theories can be unequivocally rejected, none of them are satisfactory. The main problem raised by these experiments are:

1. The unexpected effects of sentence predictability. Contrary to what might be expected (Rubenstein & Aborn, 1958; Slamecka, 1964), unpredictable sentences were learned better than predictable ones.

2. The unexpected differences between word change distractors and the cues, in particular the significance of the object in memory, and the relative unimportance of the verb.

3. The relation between LTM and STM, especially as exhibited in the changing effects of predictability and word position shift distractors over the retention interval. One cannot have a meaningful theory of

memory without time as a variable, and none of the four theories have bothered with it. The absence of the time variable makes the theories inherently untestable because each retention interval will have a different pattern of empirical findings (more or less) for comparison with the theory.

4. The relation between syntactic and semantic variables, in particular between Prd on the one hand and Yd on the other.

It seems as though learning of the sentences in these experiments could have taken place as follows. On being confronted with a sentence, a S transfers it to some STM where he resolves it into a surface structure tree. This is a relatively non-controversial assumption as there is plenty of evidence that surface structure is important in sentence perception (Neisser, 1967), as well as in sentence learning. The S then recodes the sentence by abstracting the meaning bearing elements of the sentence and linking them together so that the relationship between is preserved. This may be done via an image, or an NLM, but the basic requirement is that the elements and the relationship between them be encoded. For example, consider the following sentence:

(12) The muscular spinster defenestrated the amorous
lodger on to the blushing roses below.

There are two key images: the first is of a muscular spinster defenestrating, and the second is of the amorous lodger on his way from the window to the roses. It may be noted that the least important word is the verb, as it is implied in both the subject's and the object's image. The next most important is the locative (the adverb of place), since it is partly implied in the 'moving lodger' image.

It is of course, a purely speculative notion, but it does not seem implausible that an S, once he has resolved the tree of a sentence, breaks it into its most important semantic components, and then builds on them in some way. If he does act in this way, the factors affecting his learning rate would be the complexity of the tree, and the learnability of the key words (and their links) of the sentence, which would depend on well established variables (in other contexts) like frequency, concreteness and vividness.

In a recognition task an S presumably compares the input to his memory trace, matching first on the basis of well remembered features, and then checking the less important features of the sentence. In a cued recall task, the S presumably seeks a match for the stimulus, and then generates a sentence basedon the items which are associatively linked with the matching trace. Of course, the recognitions and cued recall situations are vastly more complicated than this, but these remarks are sufficient for our present purpose.

A model such as the one described might make the following predictions. Since the surface structure is most important at acquisition, but may be discarded once the core of a sentence has been committed to memory, then word positions shifts should be relatively more important than word changes for immediate retention than 48 hours later. This was the case. Because their surface structure trees are more easily decoded, simple sentences should be learned more easily than complex ones, and this expectation was confirmed. Since unpredictable sentences evoke unusual and rather vivid images, their components are linked more readily together (a kind of Von Restoroff effect), and so they should be learned more easily. Such as the case.

Since there was considerable time pressure on the Ss, it might be predicted that in cases where they had to decode a complex tree, as well as learn unpredictable relationships, performance would suffer. There was some evidence to support this prediction, particularly in the second experiment, but it was not strong.

The detection of new sentences and the differences between the various cues were closely related. The object was an easily detected distractor and an excellent cue, whereas the verb was neither. This finding follows naturally from the model described above. The two key components are the subject and the object. These, then, are committed first to memory, followed by the adverbial phrase, as an adjunct to the object. The verb, being implied by both the subject and the object, is, in a sense, redundant, so it is added last, if at all. Hence the verb is detected very poorly. The results of Sachs (1967) are relevant here. She found that word position changes were well detected at first, but performance rapidly declined, as in the present case. She also found that passivisation was poorly detected, a result she was not able to explain. The passive transformation, of course, operates on the verb, and since the verb is poorly stored anyway, it is not surprising that transformations on it are not detected. When she changed the subject-object relationship, detection was very good, an outcome consistent with the present results.

In sum, memory for sentences seems to depend to a considerable extent on variables well established in verbal learning such as vividness, concreteness, and STM, as well as the complexity of structural relations between individual words. Using simple measures like these, it seems possible to account for a reasonable amount of the variance in sentence acquisition.

REFERENCES

Aborn, M. & Rubenstein, H. Perception of contextually dependent word probabilities. *American Journal of Psychology*, 1958, 71, 420-422.

Aborn, M., Rubenstein, H. & Sterling, T.D. Sources of contextual constraint upon words in sentences. *Journal of experimental Psychology*, 1959, 57, 171-180.

Adams, J.A. *Human Memory*. New York: McGraw-Hill, 1967.

Bahrick, H.P. Methods of measuring retention. In E.D. Bilodeau (Ed). *Acquisition of skill*. New York: Academic Press, 1966, Pp. 35-360.

Bartlett, F.C. *Remembering: A study in experimental and social psychology*. London: C.U.P., 1932.

Battig, W.F. Procedural problems in paired-associate learning research. *Psychonomic Science Monograph Supplements*, 1965, 1, No. 1, 1-12.

Battig, W.F. Paired associate learning. In Dixon, T.R. & Horton, D.L. (Eds). *Verbal Behaviour and General Behaviour Theory*. Englewood Cliffs, New Jersey: Prentice Hall, 1968, Pp. 146-171.

Becker, S.W. & Carroll, Jean. The effects of high and low sentence contingency on learning and attitudes. *Language & Speech*, 1963, 6, 46-52.

Becker, S.W., Bavelas, A., & Braden, Marcia. An index to measure contingency of English sentences. *Language and Speech*, 1961, 4, 139-145.

Bever, T.G. Associations to stimulus-response theories of language. In Dixon, T.R. & Horton, D.L. (Eds). *Verbal Behaviour and General Behaviour Theory*. Englewood Cliffs, New Jersey: Prentice-Hall, 1968, Pp. 478-494.

Bitzer, D.L., Hicks, B., Johnson, R., & Lyman, E.R. The PLATO Systems: Current Research & Development. *IEEE Transactions on Human Factors in Electronics*, 1967, 8, 64-70.

Blumenthal, A.L. Promoted recall of sentences. Journal of verbal learning and verbal behaviour, 1967, 6, 203-206.

Blumenthal, A.L., & Boakes, R. Prompted recall of sentences. Journal of verbal learning and verbal behaviour, 1967, 6, 674-676.

Bregman, A.S. & Strasberg, Rhona. Memory for the syntactic form of sentences. Paper read to the Eastern Psychological Association Meeting, 1967.

Chomsky, A.N. Syntactic structures. The Hague: Mouton, 1957.

Chomsky, A.N. A review of Verbal Behaviour, by B.F. Skinner (Appleton-Century Crofts, New York, 1957). Language, 1959, 35, 26-58.

Chomsky, A.N. Aspects of the theory of snytax. Cambridge: MIT Press, 1965.

Chomsky, A.N. Topics in the theory of generative grammar. In T.A. Sebeok(Ed). Current Trends, in Linguistics. Vol. III. The Hague: Mouton & Co., 1966, 1-58.

Clark, K.B. Some factors influencing the remembering of prose material. Archives of Psychology, 1940, N. 253.

Clifton, C., Kurcz, Ida & Jenkins, J.J. Grammatical relations as determinants of sentence similarity. Journal of verbal learning and verbal behaviour, 1965, 4, 112-117.

Clifton, C. & Odom, Penelope. Similarity relations among certain English sentence constructions. Psychological Monographs, 1966, 80, No. 5 (Whole No. 613).

Cofer, C.N. A comparison of logical and verbatim learning of prose passage of different lengths. American Journal of Psychology, 1941, 54, 1-20.

Cofer, C.N. Some evidence for coding processes derived from clustering in free recall, *Journal of verbal learning and verbal behaviour*, 1966, 5, 188-192.

Cofer, C.N. Conditions for the use of verbal associations. *Psychological Bulletin*, 1967, 68, 1-12.

Coleman, E.B. Approximation to English: some comments on the method. *American Journal of Psychology*, 1963, 76, 239-247.

Crawford, June, Hunt, E.B. & Peak, G. Inverse forgetting in short-term memory. *Journal of experimental Psychology*, 1966, 72, 415-422.

Dale, E. & Chall, J.S. A formula for predicting readability. *Education Research Bulletin*, S.U., 1948, 27, 11-20 & 37-54.

Deese, J. On the structure of associations. *Psychological Review*, 1962, 69, 161-175.

Deese, J. *The structure of associations in language and thought*. Baltimore: The Johns Hopkins Press, 1965.

Ebbinghaus, H. *Uber das Gedachtnis*. Leipzig, 1885.

Epstein, W. The influence of syntactical structure on learning. *American Journal of Psychology*, 1961, 74, 80-85.

Epstein, W. A further study of the influence of syntactical structure on learning. *American Journal of Psychology*, 1962, 75, 121-126.

Ervin-Tripp, Susan M. & Slobin, D.I. Psycholinguistics. In P.R. Farnsworth (Ed). *Annual Review of Psychology*. Palo Alto: Annual Review Inc., 1966, 435-474.

Fillenbaum, S., Jones, L.V. & Rapoport, A. The predictability of words and their grammatical classes as a function of rate of deletion from a speech transcript. *Journal of verbal learning and verbal behaviour*, 1963, 2, 186-194.

Flesch, R.F. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

Fodor, J.A. & Bever, T.G. The psychological reality of linguistic segments. Journal of verbal learning and verbal behaviour, 1965, 4, 414-420.

Fodor, J.A. &Garrett, M. Some syntactic determinants of sentential complexity, Perception & Psychophysics, 1967, 2, 289-296.

Forster, K.I. The effects of syntactic structure on non-ordered recall. Journal of verbal learning and verbal behaviour, 1966, 5, 292-297(b).

Garrett, M., Bever, T.G. & Fodor, J.A. The active use of grammar in speech perception. Perception and Psychophysics, 1966, 1, 30-32.

Glanzer, M. Grammatical category: A rote learning and word association analysis. Journal of verbal learning and verbal behaviour, 1962, 1, 31-41.

Goldman Eisler, F. Speech production and the predictability of words in context. Quarterly Journal of experimental Psychology, 1958, 10, 96-106(a).

Goldman-Eisler, F. The predictability of words in context and the length of pauses in speech. Language and Speech, 1958, 1, 226-231(b).

Goldman-Eisler, F. Hesitation and information in speech. In C. Cherry (Ed). Information Theory-Fourth London Symposium. Washington: Butterworth, 1961, 162-174(a).

Goldman-Eisler, F. A comparative study of two hesitation phenomena. Language and Speech, 1961, 4, 18-26(b).

Gomulicki, B.R. Recall as an abstractive process. Acta Psychologica, 1956, 12, 77-94.

Gough, P.B. Grammatical transformations and speed of understanding, Journal of verbal learning and verbal behaviour, 1965, 4, 107-111.

Hake, H.W. & Rodwan, A.S. Perception and recognition in experimental methods. In Experimental Methods and Instrumentation in Psychology, J.B. Sidowski, (Ed). New York: McGraw-Hill, 1966, 331-381.

Hall, J.F. Retroactive inhibition in meaningful material. Journal of educational Psychology, 1955, 46, 47-52.

Hart, J.T. Memory and the feeling-of-knowing experience. Journal of educational Psychology, 1965, 56, 208-218.

Hart, J.T. Second-try recall, recognition, and the memory-monitoring process. Journal of educational Psychology, 1967, 58, 193-197.

Henderson, E.N. Memory for connected trains of thought. Psychological Monographs, 1907, No. 23.

Howes, D. & Osgood, C.E. On the combination of associative probabilities in linguistic contexts. American Journal of Psychology, 1954, 67, 241-259.

Hull, C.L. Knowledge and purpose as habit mechanisms. Psychological Review, 1930, 37, 511-525.

Humphrey, G. Thinking: an introduction to its experimental psychology. New York: Wiley, 1951.

Johnson, N.F. The psychological reality of phrase-structure rules. Journal of verbal learning and verbal behaviour, 1965, 4, 469-476(a).

Johnson, N.F. Linguistic models and functional units of language behaviour. In S. Rosenberg (Ed) Directions in Psycholinguistics. New York: Macmillan, 1965(b).

Johnson, N.F. The influence of associations between elements of structural verbal responses. Journal of verbal learning & verbal behaviour, 1966, 5, 369-74(a).

Johnson, N.F. On the relationship between sentences structure and the latency in generating the sentence. Journal of verbal learning and verbal behaviour, 1966, 5, 375-389(b).

Johnson, N.F. Sequential verbal behaviour. In Dixon, T.R. & Horton, D.L. (Eds). Verbal Behaviour and General Behaviour Theory. Englewood Cliffs, N.J.: Prentice-Hall, 1968, 421-450.

Katona, G. Organizing and memorizing. New York: Columbia University Press, 1940.

Kay, H. Learning and retaining verbal material. British Journal of Psychology, 1955, 46, 81-100.

King, D.J. On the accuracy of written recall; a scaling and factor analytic study. Psychological Record, 1960, 10, 113-122.

King, D.J. Scaling the accuracy of recall of stories in the absence of objective criteria. Psychological Record, 1961, 11, 87-90.

King, D.J. & Lau, A.W. A comparison of three scaling techniques in estimating the accuracy of written recall. Journal of general Psychology, 1963, 69, 203-207.

King, D.J. & Russell, G.W. A comparison of rote and meaningful learning of connected meaningful material. Journal of verbal learning and verbal behaviour, 1966, 5, 478-483.

King, D.J. & Schultz, D.P. Additional observations on scoring the accuracy of written recall. Psychological Record, 1960, 10, 203-209.

King, D.J. & Yu, K.C. The effects of reducing the variability of length of written recalls on the rank order scale values of the recalls. Psychological Record, 1962, 12, 39-44.

Klare, C.R. Measures of the readability of written communication: an evaluation. *Journal of educational Psychology*, 1952, 43, 385-399.

Lewis, F.H. Note on the doctrine of memory traces. *Psychological Review*, 1933, 40, 90-97.

Levine, J.M. & Murphy, G. The learning and forgetting of controversial material. *Journal of abnormal and social Psychology*, 1941, 38, 507-517.

Levitt, E.E. A methodological study of the preparation of connected verbal stimuli for quantitative memory experiments. *Journal of experimental Psychology*, 1956, 52, 33-38.

Lounsbury, F.G. Sequential psycholinguistics. In Osgood, C.E. & Seboek, T. A. (Eds). Psycholinguistics: a survey of theory and research problems. *Journal of abnormal and social Psychology*, 1954, 49, suppl., 93-125.

Lyon, D.O. The relation of quickness of learning to retentiveness. *Archives of Psychology*, 1916, Whole No. 34.

Lyon, D.O. *Memory and the learning process*. Baltimore: Warwick & York, 1917.

McNulty, J.A. The measurement of "adopted chunks" in free recall learning. *Psychonomic Science*, 1966, 4, 71-72.

Maclay, H. & Osgood, C.E. Hesitation phenomena in spontaneous English speech. *Word*, 1959, 15, 19-44.

Mandler, G.M. Response factors in human learning. *Psychological Review*, 1954 61, 235-44.

Mandler, G.M. From association to structure. *Psychological Review*, 1962, 69, 415-427.

Mandler, G.M. Organization and Memory. In K.W. Spence & Janet T. Spence (Eds). *The Psychology of Learning and Motivation*. New York: Academic Press, 1967.

Mandler, G.M. Verbal Learning. In New Directions in Psychology III. New York: Holt, Rinehart & Winston, Inc., 1967.

Mandler, G.M. Association and organization: factors, fancies and theories. In Dixon, T.R. & Horton, D.L. Verbal Behaviour and General Behaviour Theory. Englewood Cliffs, N.J.: Prentice-Hall, 1968, 109-120.

Mandler, G.M. & Mandler, Jean M. Serial position effects in sentences. Journal of verbal learning and verbal behaviour, 1963, 3, 195-202.

Mark, M.R. & Jacks, O. Verbal context and memory span for meaningful material. American Journal of Psychology, 1952, 65, 298-300.

Marks, L.E. & Miller, G.A. The role of semantic and syntactic constraints in the memorization of English sentences. Journal of verbal learning and verbal behaviour, 1964, 3, 1-5.

Marshall, G.R. & Cofer, C.N. Associated indices as measures of word relatedness: a summary and a comparison of ten methods. Journal of verbal learning and verbal behaviour, 1963, 1, 408-421.

Martin, E. & Roberts, K.H. Grammatical factors in sentence retention. Journal of verbal learning and verbal behaviour, 1966, 5, 211-213.

Martin, E. & Roberts, K.H. Sentence length and sentence retention in the free learning situation. Psychonomic Science, 1967, 8, 535-536.

Martin, E., Roberts, K.H. & Collins A.M. Short-term memory for sentences. (In Press, 1968).

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. Journal of verbal learning and verbal behaviour, 1963, 2, 346-351.

Miller, G.A. The magical number seven plus or minus two: some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Miller, G.A. Some psychological studies of grammar. *American Psychologist*, 1962, 17, 748-762.

Miller, G.A. & Chomsky, A.N. Finite models of language users. In Luce, R.D., Bush, R.R. & Galanter E. (Eds) *Handbook of mathematical psychology*, Vol. II. Wiley: 1967.

Miller, G.A., Galanter, E. & Bribam, K. *Plans and the structure of behaviour*. New York: Henry Holt, 1960.

Miller, G.A. & Isard S. Some perceptual consequences of linguistic rules. *Journal of verbal learning and verbal behaviour*, 1963, 2, 217-228.

Miller, G.A. & Isard, S. Free recall of self-embedded English sentences. *Information & Control*, 1964, 7, 292-303.

Miller, G.A. & McKean, Kathleen O. A chronometric study of some relations between sentences. *Quarterly Journal of experimental Psychology*, 1964, 16, 297-308.

Miller, G.A. & Selfridge, J.A. Verbal context and the recall of meaningful material. American Journal of Psychology, 1950, 63, 176-185.

Montague, W.E. & Wearing, A.J. The complexity of natural language mediators and its relation to paired-associate learning. *Psychonomic Science*, 1967, 7, 135-136.

Montague, W.E., Adams, J.A., & Kiess, H.O. Forgetting and natural language mediation. *Journal of experimental Psychology*, 1966, 72, 829-833.

Murdock, B.B. Signal detection theory and short-term memory. *Journal of experimental Psychology*, 1965, 70, 443-447.

Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1967.

Noble, C.E. An analysis of meaning. Psychological Review, 1952, 59, 421-430.

Norman, D.A. & Wickelgren, W.A. Short-term recognition memory for single digits and pairs of digits. Journal of experimental Psychology, 1965, 70, 479-489.

Osgood, C.E. Method and Theory in Experimental Psychology, New York: O.U.P. 1953.

Osgood, C.E. On understanding and creating sentences. American Psychologist, 1963, 18, 735-751.

Osgood, C.E. Towards a wedding of insufficiencies. In Dixon, T.R. & Horton, D.L. Verbal Behaviour and General Behaviour Theory. Englewood Cliffs, N.J.: Prentice-Hall, 1968, 495-520.

Pappas, B.A. & Suboski M.D. Recognition operating characteristics as a function of prior recall confidence. Psychonomic Science, 1966, 6, 83-84.

Pompi, K.F. & Lachman, R. Surrogate processes in the short-term retention of connected discourses. Journal of experimental Psychology, 1967, 75, 143-150.

Postman, L. & Adams, Pauline. Studies on incidental learning. VII. The effects of contextual determination. Journal of experimental Psychology, 1960, 59, 153-164.

Richardson, Patricia, & Voss, J.F. Replication Report: Verbal context and the recall of meaningful material. Journal of experimental Psychology, 1960. 60, 417-418.

Rohwer, W.D. Constraint, syntax, and meaning in paired-associate learning. Journal of verbal learning and verbal behaviour, 1966, 5, 541-547.

Rohwer, W.D. & Lynch, S. Semantic constraint in paired-associate learning. Journal of educational Psychology, 1966, 57, 27-78.

Rohwer, W.D., Shuell, T.J. & Levin, J.R. Context effects in the initial storing and retrieval of noun pairs. Journal of verbal learning and verbal behaviour, 1967, 6, 796-801.

Rosenberg, S. Recall of sentences as a function of syntactic and associative habit. Journal of verbal learning and verbal behaviour, 1960, 5, 392-396.

Rosenberg, S. Grammatical and associative factors in sentence recall: a preliminary report in Lane and Zale (Eds) Studies in Language and Language Behaviour, IV, Ann Arbor, Center for Research on Language and Language Behaviour, University of Michigan, 1967, 292-301.

Rosenberg, S. The relation between association and syntax in sentence production. In Catford, J.C. (Ed). Studies in Language and Language Behaviour, V, Sept. 1, 1967, Center for Research on Language and Language Behaviour, Ann Arbor, University of Michigan.

Rubenstein, H. & Aborn, M. Learning, prediction and readability. Journal of Applied Psychology, 1958, 42, 28-32.

Russell,W.A. & Jenkins, J.J. The complete Minnesota norms for responses to words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, University of Minnesota, Contract N8ONR-66216, Office of Naval Research, 1954.

Sachs, Jacqueline S. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442.

Salzinger, K. & Eckerman, Carol. Grammar and the recall of chains of verbal responses. Journal of verbal learning & verbal behaviour, 1967, 6,232-239.

Salzinger, K., Portnoy, Stephanie, & Feldman, R.S. The effect of order of approximation to the statistical structure of English on the emission of verbal responses. *Journal of experimental Psychology*, 1962, 64, 52-57.

Savin, H.B. & Perchoneck, E. Grammatical structure and the immediate recall of English sentences. *Journal of verbal learning and verbal* behaviour, 1965, 4, 348-353.

Sharp, H.C. Effect of contextual constraint upon recall of verbal passages. *American Journal of Psychology*, 1958, 71, 568-572.

Shepard, R.N. Production of constrained associates, and the informational uncertainty of the constraint. *American Journal of Psychology*, 1963, 76, 218-228.

Slamecka, N.J. Studies of retention of connected discourse. *American Journal of Psychology*, 1959, 72, 409-416.

Slamecka, N.J. Retroactive inhibition of connected discourse as a function of practice level. *Journal of experimental Psychology*, 1961, 59, 104-108.

Slamecka, N.J. Retention of connected discourse as a function of interpolated learning. *Journal of experimental Psychology*, 1962, 63, 480-486.

Slamecka, N.J. An inquiry into the doctrine of remote associations. *Psychological Review*, 1964, 71, 61-76(a).

Slamecka, N.J. Acquisition and retention of connected restraint. *Journal of experimental Psychology*, 1964, 68, 330-333(b).

Slamecka, N.J. and Ceraso, J. Retroactive and proactive inhibition of verbal learning. *Psychelgical Bulletin*, 1960, 57, 449-475.

Slobin, D.I. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of verbal learning and verbal behaviour*, 1966, 5, 219-227..

Staats, A.W. & Staats, C.K. *Complex Human Behaviour*, New York: Holt, Rinehart & Winston, 1966.

Suci, G.J. The validity of pause as an index of units in language. *Journal of verbal learning and verbal behaviour*, 1967, 6, 26-32.

Tannenbau, P.H., Williams, F. & Hillier, Carolyn, S. Word predictability in the environment of hesitations. *Journal of verbal learning and verbal behaviour*, 1965,4, 134-140.

Taylor, W.L. "Cloze procedure": a new tool for measuring readability. *Journalism Quarterly*, 1953, 30, 415-433.

Taylor, W.L. Recent developments in the use of the "Cloze procedure". *Journalism Quarterly*, 1956, 33, 42-48.

Thorndike, E.L. & Lorge, I. *The Teachers word book of 30,000 words*. New York: Teachers College, 1944.

Treisman, Anne M. Verbal responses and contextual constraints in language. *Journal of verbal learning and verbal behaviour*, 1965, 4, 118-128.

Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychological Review*. 1962, 69, 344-354.

Tulving, E. Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychological Review*, 1964, 71, 219-237.

Tulving, E. Theoretical issues in free recall. In Dixon, T.R. & Horton, D.L. (Eds). *Verbal Behaviour and General Behaviour Theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1968, 2-36.

Tulving, E. & Patkau, J.E. Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology*, 1962, 16, 83-95.

Underwood, B.J. *Experimental Psychology*. New York: Appleton-Century-Crofts, 1965.

Underwood, B.J. The language repertoires and some problems in verbal learning. In S. Rosenberg (Ed) *Directions in Psycholinguistics*, New York: Macmillan, 1965, 99-120.

Underwood, B.J. False recognition produced by implicit verbal responses. *Journal of experimental Psychology*, 1965, 70, 122-129(c).

Underwood, B.J. & Schultz, R.W. *Meaningfulness and verbal learning*. Chicago: Lippincott, 1960.

Wales, R.J. & Marshall, J.C. The organization of linguistic performance. In *Psycholinguistic Papers*. (Ed J. Lyons & R.J. Wales). Edinburgh. Edinburgh University Press, 1966.

Wearing, A.J. Retrieval and Adams response recognition hypothesis. In preparation.

Weinreich, U. Explorations in semantic theory. In *Current Trends in Linguistics* (Ed. T.A. Sebeok). Vol. III Hague: Mouton & Co., 1966, 397-478.

Welborn, E. & English, H.B. Logical learning and retention: a general review of experiments with meaningful verbal material. *Psychological Bulletin*, 1937, 34, 1-20.

Winograd, E. & Von Saal, W. Discriminability of association value in recognition memory. *Journal of experimental Psychology*, 1966, 72, 328-334.

Woodworth, R.S. & Schlosberg, H. *Experimental Psychology*. New York: Holt, Rinehart & Winston, 1954.

Young, R.K. Serial Learning. In Dixon, T.R. & Horton, D.L. (Eds). *Verbal Behaviour and General Behaviour Theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1968, 122-245.

Yngve, V.H. A model and an hypothesis for language structure. *Procedure for American Philosophical Society*, 1960, 104, 444-466.

APPENDIX I

This appendix contains the 64 experimental sentences which were learned by the S's in both experiments.

The tidy carpenter carefully threw the metal dish in the shallow puddle.

The icy pepsi from a crystal mug was demurely sampled by the radiant queen.

The bronzed pioneer in a hushed tone excitedly related his own history.

The keen birdwatcher soon sighted with his strong binoculars a rare owl.

At the busy airport the tedious delay was soon resented by the alien traveller.

The weak patient was slapped savagely by the medical sister with the wet towel.

The thin mimer was loudly honoured in the gigantic theatre by the lengthy applause.

The friendly physician was always greeted with courteous nods by the hospital matron.

During the brief crisis the frenzied mob was greatly stirred by its own fear.

The brown rabbit anxiously eyed the circling eagle in the clear sky.

The blind adult politely led through the misty forest the rude hooligan.

The rich director in the filthy bus sternly discussed current news.

The scared cows during the storm were kindly soothed by several boys.

On the bleak mountains the hardy peasants patiently tended their white goats.

During the hectic campaign ripe bananas were slowly munched by the confident nominee.

The dynamic orator during the open session briskly reiterated the major facts.

With a happy smile the pretty typist often inhaled an expensive cigar.

The ancient king bent regally with an impressive dignity his gnarled frame.

Foreign policies are wrongly moulded by verbal radicals in the eastern towns.

The slim acrobat with unique ability swiftly leapt into the bright light.

Into the shocked onlookers the fresh peach was deftly socked by the surly dentist

The sleeping hiker beneath the cool shelter was abruptly woken by chirping robins.

The majestic trout was gladly returned by the ravenous hunter to the wide river.

The fragile doll in one action was badly wrecked by the tall commander.

The inferior side fought bravely its tough opponents to the bitter finish.

A prized medal was graciously proffered by the chairman to the perspiring athlete.

In the local store the observant helper neatly mended the rotten stair.

The wealthy earl contentedly lived in the Negro ghetto his last years.

The militant fascist was lavishly treated with grilled veal by the University President.

The ill beggar was briefly overcome on the damp seat by sharp pains.

In the mink coat the barefoot pupil hastily took the admiral's hand.

The bank teller cheerfully carried along the bumpy road the stolen money.

The stone church was loyally supported by the small village in many ways.

In damp crannies fresh pastry was perpetually consumed by heroin users.

The massive elm in the mild weather was thinly sprayed with soft showers.

The sterm ladies drunkenly enjoyed in the temperance club the cocktail banquet.

The snarling pup with brutal pressure was roughly hit by the grim policeman.

The clever lecturer from the renowned college once forgot his precious notes.

The fiery speaker proclaimed defiantly on a chilly eve his single avewal.

The southern politician lustily yodelled many freedom chants out in the dry plain.

The gaudy shirt beneath silk robes was gaily displayed by the grave cardinal.

The tired artist in pale blue pyjamas solemnly addressed the empty senate.

The tame squirrel grasped nimbly with an easy grace a swaying bough.

The scarlet pants were proudly chosen for special trips by the dignified chauffeur.

In two days the foreign occupant was warmly accepted by the wary landlord.

The stocky felon for a few cookies silently raided the full hotel.

The cynical newsman exposed recently the secret betrayal with a rueful snort.

The grey kangaroo joyously trampled the growing barley with his clumsy tread.

On the lonely shore bearded sailors eagerly fried a tasty omelette.

From his fast jeep the county sheriff quickly drew out his trusty pistol.

The unhappy nation was grieviously torn by virulent pestilences for some decades.

The uncouth bishop was hugged passionately in the musty inn by the prim spinster.

The sleek beaver on the floating raft was greedily followed by hungry serpents.

In the waiting taxi the loud eruption suddenly ended the tense stillness.

At early daybreak some stale custard was shyly deposited by the dark postman.

The alert author into the whirling tornado merrily drove an old truck.

The verbose host accidentally shook a dusty mop in the janitor's face.

The weary pony was gently urged along the stony paths by the bald friar.

The praying native was cruelly killed in the grimy hut by the Christian pacifist.

The shaking minister nervously placed the loaded weapon on the tiled roof.

With much care the small package was lovingly hidden by his aunt.

The ugly boss was grandly entertained by the poor cleaner with Swiss liqueur.

The raucous jury was boldly given by the miserable prisoner a free whisky.

The freckled youth sheepishly washed his grubby hands in the clear water.

APPENDIX 2

Instructions given to the subjects in experiments 1 & 2.

Slide 1.

WELCOME TO PLATO

You are about to take part in an
experiment on memory for sentences.

Press the space bar for
more instructions.

Slide 2.

Each sentence will be shown for 25 seconds on this screen. During that time you must write down the sentence in the booklet on your right.

To Continue,

Press the Space bar

Slide 3.

As you write down each sentence, try to memorize it. Later on, you will be tested to see how effective your memory is.

Press the space bar

Slide 4.

Write quickly. Three seconds before the next sentence is going to be presented, the words "three seconds left" will appear towards the buttom of the screen. This is a signal to hurry.

Press the space bar

Slide 5.

To ensure that you understand the procedure, you will be given a set of practice sentences to learn.

Before learning the sentences, you may want to review the instructions. If so, press the key marked B, and you will back up through the instructions.

When you are satisfied that you know what to do, press the space bar, and the sentences will be displayed, one at a time, for 25 secs. each.

Press the space bar.

Slide 6.

Write down and learn this
sentence as quickly as possible.

Slide 7.

That completes the list of sentences.

Slide 8.

If you have any questions at all, raise your hand and call an experimenter.

If you feel quite confident about the procedure, press the space bar to continue.

Slide 9.

You will now be presented with
68 sentences to write down and learn.

Slide 10.

We come now to the retention test. In this test sentences will be presented to you one at a time.

Some of these sentences will be from the list that you learned, others will be new, although they will be similar to the "old" sentences in the list that you learned.

If the sentence shown is from the list that you learned, press the key marked "old". If the sentence is new (that is, it was not on the learning list), press the key marked "new".

Press the space bar.

Slide 11.

You will only have about 7 seconds in which to indicate whether or not the sentence is old or new, and to make your confidence judgment.

Three seconds before the next sentence is presented, the words "three seconds left" will appear towards the bottom of the screen.

Press the space bar

Slide 12.

After having pressed the "old" or the "new" key, the next task is to indicate how confident you are that you pressed the correct key.

To make this judgment, press one of the keys marked 1, 2, 3, 4, 5, or 6, where:

(Press the space bar to see what these numbers mean).

Slide 13.

1 = completely (100%) sure that the response was correct

2 = 80% sure that the response was correct.

3 = 60% sure that the response was correct.

4 = 40% sure that the response was correct.

5 = 20% sure that the response was correct.

6 = no idea (0% certainty) whether or not the response was correct.

Press the space bar

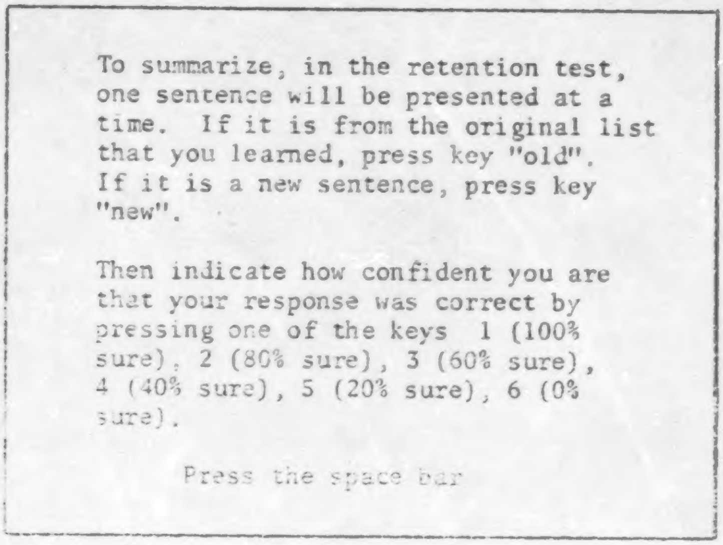
Slide 14.

Three points to remember:

1. Press the "old" or "new" button as soon as you can, but do not forget that you are allowed only one response.
2. Always press a key. If you do not know the answer, then guess.
3. When making your confidence judgments, use as many of the six categories as you can.

When you have these points fixed in your memory, press the space bar.

Slide 15.



To summarize, in the retention test, one sentence will be presented at a time. If it is from the original list that you learned, press key "old". If it is a new sentence, press key "new".

Then indicate how confident you are that your response was correct by pressing one of the keys 1 (100% sure), 2 (80% sure), 3 (60% sure), 4 (40% sure), 5 (20% sure), 6 (0% sure).

Press the space bar

Slide 16.

You can review any of the instructions by pressing key B.

To ensure that you understand the procedure you will be given a short practice test.

If you are confident that you understand what to do, press the space bar.

Slide 17.

Indicate whether this sentence is "old" or "new" by pressing key "old" or key "new".

Indicate how confident you are that your response is correct by pressing one of the keys:

1	2	3	4	5	6
(100% sure)	(80% sure)	(60% sure)	(40% sure)	(20% sure)	(0% sure)

Slide 18.

We come now to the retention test. In this test two cue words chosen from one of the sentence that was in the list that you learned will be presented to you, one pair of words at a time.

Your task is to write down as much of the remainder of the sentence as you can remember.

Press the space bar.

Slide 19.

For example, if the original sentence was:

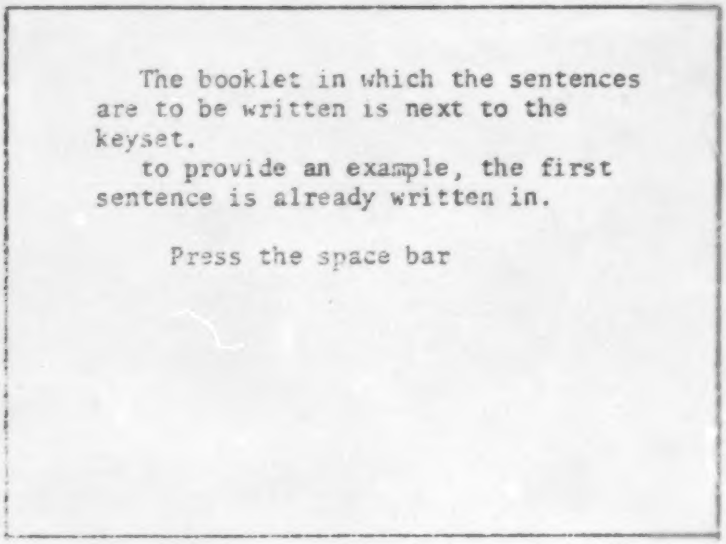
"only a knave can move
the levers of political
power",

the two cue words might be,

"political power"

Press the space bar

Slide 20.



The booklet in which the sentences are to be written is next to the keyset.

to provide an example, the first sentence is already written in.

Press the space bar

Slide 21.

If you want to review any of the instructions, press key B, to back up.

When you feel that you understand the procedure, press next.

Slide 22.

First, there is a short practice test. You will only get 35 seconds, so you will have to write quickly.

Three seconds before the next pair of cue words is presented, the words "three seconds left" will appear towards the bottom of the screen.

Press the space bar

Slide 23.

CUE WORDS

Write out the rest of
the sentence as quickly
as possible.

Slide 24.

You will now be presented
with 68 pairs of cue words,
one pair at a time.

APPENDIX 3

Raw Data

VITA

Alexander James Wearing was born near Weetulta, South Australia, on March 20th, 1940. He was schooled in various places in South Australia, finally graduating in 1963 from the University of Adelaide, South Australia. In 1964, Mr. Wearing became a graduate student at the University of Illinois in Urbana. In 1968 he joined the faculty of Yale University.